

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 22:17:24 ; Search time 183 Seconds  
(without alignments)  
48.020 Million cell updates/sec

Title: US-10-797-821-22

Perfect score: 116  
Sequence: 1 QEANYAGNSIGNYRGWNP 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*
- 9: Geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 116   | 100.0       | 20     | 7     | ADD93642    |
| 2          | 116   | 100.0       | 20     | 9     | AD37265     |
| 3          | 116   | 100.0       | 431    | 7     | ADD93649    |
| 4          | 116   | 100.0       | 431    | 7     | ADD93650    |
| 5          | 116   | 100.0       | 431    | 7     | ADD93653    |
| 6          | 116   | 100.0       | 431    | 9     | AD37272     |
| 7          | 116   | 100.0       | 431    | 9     | AD37273     |
| 8          | 116   | 100.0       | 431    | 9     | AD37276     |
| 9          | 116   | 100.0       | 431    | 9     | AE91500     |
| 10         | 116   | 100.0       | 432    | 7     | ADD93651    |
| 11         | 116   | 100.0       | 432    | 7     | ADD93652    |
| 12         | 116   | 100.0       | 432    | 9     | AD37274     |
| 13         | 116   | 100.0       | 432    | 9     | AD37275     |
| 14         | 96    | 82.8        | 132    | 2     | AAW60944    |
| 15         | 96    | 82.8        | 392    | 6     | ABU02747    |
| 16         | 96    | 82.8        | 392    | 8     | ADK47859    |
| 17         | 96    | 82.8        | 392    | 8     | ADT50227    |
| 18         | 96    | 82.8        | 392    | 8     | ADT50226    |
| 19         | 96    | 82.8        | 392    | 8     | ADT50165    |
| 20         | 96    | 82.8        | 399    | 8     | ADR94595    |
| 21         | 96    | 82.8        | 399    | 9     | AEA58465    |
| 22         | 89    | 76.7        | 395    | 5     | ABP25918    |
| 23         | 89    | 76.7        | 447    | 5     | ABP29684    |
| 24         | 89    | 76.7        | 447    | 8     | ADU69524    |

|    |    |      |     |   |          |          |           |
|----|----|------|-----|---|----------|----------|-----------|
| 25 | 89 | 76.7 | 447 | 8 | ADV88392 | Adv88392 | Streptoco |
| 26 | 89 | 76.7 | 447 | 8 | ADV81808 | Adv81808 | Streptoco |
| 27 | 89 | 76.7 | 447 | 8 | ADV79645 | Adv79645 | Streptoco |
| 28 | 84 | 72.4 | 165 | 5 | ABP29842 | Abp29842 | Streptoco |
| 29 | 84 | 72.4 | 169 | 5 | ABP27864 | Abp27864 | Streptoco |
| 30 | 84 | 72.4 | 169 | 8 | ADU69652 | Adu69652 | S agalact |
| 31 | 84 | 72.4 | 169 | 8 | ADV88318 | Adv88318 | Streptoco |
| 32 | 84 | 72.4 | 169 | 8 | ADV81739 | Adv81739 | Streptoco |
| 33 | 84 | 72.4 | 169 | 8 | ADV79571 | Adv79571 | Streptoco |
| 34 | 77 | 66.4 | 398 | 5 | ABP25919 | Abp25919 | Streptoco |
| 35 | 77 | 66.4 | 398 | 8 | ADR83884 | Adr83884 | S. pyogen |
| 36 | 77 | 66.4 | 398 | 9 | AE91542  | Aeb91542 | Microbial |
| 37 | 74 | 63.8 | 544 | 9 | AE91648  | Aeb91648 | Microbial |
| 38 | 70 | 60.3 | 211 | 9 | AE91642  | Aeb91642 | Microbial |
| 39 | 50 | 43.1 | 133 | 4 | AAU19320 | Aau19320 | Human G p |
| 40 | 50 | 43.1 | 510 | 9 | ARM94545 | Arm94545 | M. xanthu |
| 41 | 49 | 42.2 | 123 | 3 | AAV87588 | Aav87588 | Aminopept |
| 42 | 49 | 42.2 | 308 | 3 | AAV87586 | Aav87586 | Aeromonas |
| 43 | 49 | 42.2 | 407 | 3 | AAV87585 | Aav87585 | Aeromonas |
| 44 | 49 | 42.2 | 491 | 3 | AAV87584 | Aav87584 | Aeromonas |
| 45 | 49 | 42.2 | 590 | 3 | AAV87583 | Aav87583 | Aeromonas |

ALIGNMENTS

RESULT 1  
ADD93642  
ID ADD93642 standard; peptide; 20 AA.  
XX  
AC ADD93642;  
XX  
XX 29-JAN-2004 (first entry)  
XX  
DE Streptococcus mutans glucan binding protein-B peptide fragment.  
XX  
KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.  
XX  
OS Streptococcus mutans.  
XX  
PN WO2003075845-A2.  
XX  
PD 18-SEP-2003.  
XX  
PF 07-MAR-2003; 2003WO-US006962.  
XX  
PR 07-MAR-2002; 2002US-0363209P.  
PR 08-AUG-2002; 2002US-0402483P.  
XX  
XX (FORS-) FORSYTH INST.  
PA Smith DJ, Taubman MA;  
PI  
XX  
XX  
DR WPI; 2003-845091/78.  
XX  
PT Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.  
XX  
PS Claim 11; Page 10; 49pp; English.  
XX  
XX  
CC The present sequence is that of a peptide comprising amino acid residues 403-422 of the glucan binding protein-B (GbpB) of Streptococcus mutans strain SK32 ADD93649. The peptide binds to a major histocompatibility complex (MHC) class II protein. It was identified as a potential B cell epitope using a matrix-based algorithm for epitope prediction, which was used to search the primary amino acid sequence of GbpB for known MHC class II binding motifs. The peptide can be used in immunogenic compositions and subunit vaccines for dental caries. These comprise MHC class II protein-binding GbpB peptides covalently linked with peptide subunits (preferably from the catalytic domain) of a glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Dieptopic or multiptopic polypeptides can be

CC prepared synthetically or by recombinant DNA technology. Antibodies  
 CC raised against MHC class II binding fragments of GbpB can be used in  
 CC passive immunisation.

SQ Sequence 20 AA;  
 Query Match 100.0%; Score 116; DB 7; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-11;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEANYAGNQSIGNYRGWFNP 20  
 |||||  
 Db 1 QEANYAGNQSIGNYRGWFNP 20  
 |||||

RESULT 2  
 ADX37265  
 ID ADX37265 standard; peptide; 20 AA.  
 XX  
 AC ADX37265;  
 XX  
 DT 21-APR-2005 (first entry)  
 XX  
 DE Streptococcus mutant glucan binding protein B peptide #22.  
 XX  
 KW immunogenicity; immune stimulation; glucan binding protein-B;  
 KW microparticle; major histocompatibility complex; tooth disease.  
 XX  
 OS Streptococcus mutans.  
 XX  
 PN US2005031633-A1.  
 XX  
 PD 10-FEB-2005.

PF 09-MAR-2004; 2004US-00797821.  
 XX  
 PR 13-APR-1998; 98US-0081550P.  
 PR 08-JAN-1999; 99US-0115142P.  
 PR 12-APR-1999; 99US-00290049.  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 PR 07-MAR-2003; 2003US-00383930.  
 PA (SMIT/) SMITH D J.  
 PA (TAUB/) TAUBMAN M A.  
 XX  
 PI Smith DJ, Taubman MA;  
 XX  
 DR WPI; 2005-151644/16.  
 XX  
 PT New composition comprising a fragment of a glucan binding protein-B  
 PT (GbpB) that binds to MHC class II protein, and a biocompatible  
 PT microparticle, useful for producing an antibody (claimed) for immunizing  
 PT mammals against dental caries.

PS Claim 4; SEQ ID NO 22; 73pp; English.  
 XX  
 CC The invention relates to a composition comprising a fragment of a glucan  
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
 CC fragment binds to a major histocompatibility complex (MHC) class II  
 CC protein. The composition is useful for producing an antibody for  
 CC immunizing mammals against dental caries. This sequence corresponds to a  
 CC Streptococcus mutans GbpB-derived peptide of the invention.

SQ Sequence 20 AA;  
 Query Match 100.0%; Score 116; DB 9; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-11;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEANYAGNQSIGNYRGWFNP 20  
 |||||  
 Db 1 QEANYAGNQSIGNYRGWFNP 20  
 |||||

RESULT 3  
 ADD93649  
 ID ADD93649 standard; protein; 431 AA.  
 XX  
 AC ADD93649;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Streptococcus mutans glucan binding protein-B.  
 KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.  
 XX  
 OS Streptococcus mutans.  
 XX  
 PH Key Location/Qualifiers  
 FT Region 6..25  
 FT /note= "HLA-binding peptide"  
 FT Region 16..35  
 FT /note= "HLA-binding peptide"  
 FT Region 33..52  
 FT /note= "HLA-binding peptide"  
 FT Region 37..56  
 FT /note= "HLA-binding peptide"  
 FT Region 48..67  
 FT /note= "HLA-binding peptide"  
 FT Region 52..71  
 FT /note= "HLA-binding peptide"  
 FT Region 88..107  
 FT /note= "HLA-binding peptide"  
 FT Region 113..132  
 FT /note= "HLA-binding peptide"  
 FT Region 117..136  
 FT /note= "HLA-binding peptide"  
 FT Region 137..156  
 FT /note= "HLA-binding peptide"  
 FT Region 174..193  
 FT /note= "HLA-binding peptide"  
 FT Region 194..213  
 FT /note= "HLA-binding peptide"  
 FT Region 214..233  
 FT /note= "HLA-binding peptide"  
 FT Region 248..267  
 FT /note= "HLA-binding peptide"  
 FT Region 289..308  
 FT /note= "HLA-binding peptide"  
 FT Region 306..325  
 FT /note= "HLA-binding peptide"  
 FT Region 311..330  
 FT /note= "HLA-binding peptide"  
 FT Region 349..368  
 FT /note= "HLA-binding peptide"  
 FT Region 365..384  
 FT /note= "HLA-binding peptide"  
 FT Region 383..402  
 FT /note= "HLA-binding peptide"  
 FT Region 403..422  
 FT /note= "HLA-binding peptide"  
 XX  
 WO2003075845-A2.  
 PD 18-SEP-2003.  
 XX  
 PF 07-MAR-2003; 2003WO-US006962.  
 XX  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 XX  
 PA (FORS-) FORSYTH INST.  
 XX  
 PI Smith DJ, Taubman MA;  
 XX







PF 09-MAR-2004; 2004US-00797821.  
 XX 13-APR-1998; 98US-0081550P.  
 PR 08-JAN-1999; 99US-0115142P.  
 PR 12-APR-1999; 99US-00290049.  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 PR 07-MAR-2003; 2003US-00383930.  
 XX (SMIT/) SMITH D J.  
 PA (TAUB/) TAUBMAN M A.  
 XX Smith DJ, Taubman MA;  
 PI WPI; 2005-151644/16.  
 DR  
 XX New composition comprising a fragment of a glucan binding protein-B (GbpB) that binds to MHC class II protein, and a biocompatible microparticle, useful for producing an antibody (claimed) for immunizing mammals against dental caries.  
 PT  
 PS Claim 3; SEQ ID NO 33; 73pp; English.  
 XX The invention relates to a composition comprising a fragment of a glucan binding protein-B (GbpB) and a biocompatible microparticle, where the fragment binds to a major histocompatibility complex (MHC) class II protein. The composition is useful for producing an antibody for immunizing mammals against dental caries. This sequence corresponds to a Streptococcus mutans GbpB protein of the invention.  
 CC  
 XX Sequence 431 AA;  
 SQ  
 Query Match 100.0%; Score 116; DB 9; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-10;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QYANYAGNQSIGNRGWFNP 20  
 Db |||||  
 403 QYANYAGNQSIGNRGWFNP 422  
 RESULT 9  
 ID AEB91500  
 XX AEB91500 standard; protein; 431 AA.  
 AC AEB91500;  
 XX 20-OCT-2005 (first entry)  
 DT  
 XX Microbial pathogen adhesin protein sequence, SEQ ID NO:210.  
 DE algorithm; adhesin; pharmaceutical; vaccine; drug screening;  
 KW bordetella pertussis infection; antibacterial; pneumonia;  
 KW antiinflammatory; respiratory-gen.; gastric ulcer; antidiacer;  
 KW gastrointestinal-gen.; urinary tract infection; antimicrobial; uropathic.  
 XX Streptococcus mutans.  
 OS  
 XX WO2005076010-A2.  
 PN  
 XX 18-AUG-2005.  
 PD  
 XX 07-FEB-2005; 2005WO-IN000037.  
 PF  
 XX 06-FEB-2004; 2004IN-DE000173.  
 PR 20-JUL-2004; 2004US-0589227P.  
 PR  
 XX (COUL ) COUNCIL SCI & IND RES SOUTH AFRICA.  
 PA Sachdeva G, Kumar K, Jain P, Brahmachari SK, Ramachandran S;  
 XX WPI; 2005-597835/61.  
 DR  
 XX

PT Computational method for identifying adhesin and adhesin like molecules, comprises computing sequence-based attributes of protein sequences using neural network software and training an artificial neural network.  
 PT  
 XX Claim 16; SEQ ID NO 210; 402pp; English.  
 PS  
 XX The present invention relates to a computational method (M1) for identifying adhesin and adhesin-like proteins, by computing the sequence-based attributes of protein sequences using five attribute modules of a neural network software, training an artificial neural network (ANN) for each of the computed five attributes, and identifying the adhesin and adhesin-like proteins having probability of being an adhesin (Pad) as equal or greater than 0.51. Also claimed is a set of 274 annotated genes encoding adhesin and adhesin-like proteins, having 274 fully defined 162-1151 base pair (SEQ ID NO: 385-658) sequences; a set of 105 hypothetical genes encoding adhesin and adhesin-like proteins, having 105 fully defined 306-15876 base pairs (SEQ ID NO: 659-763); a set of 279 annotated adhesin and adhesin-like proteins, having 279 fully defined 53-3716 base pair (SEQ ID NO: 1-279); a set of 105 hypothetical adhesin and adhesin-like proteins, having 105 fully defined 106-5291 base pair (SEQ ID NO: 280-384) sequences; and a fully connected multilayer feed forward ANN (1) based on (M1). (M1) is useful for identifying adhesin and adhesin-like proteins, of therapeutic potential, and identifying and short-listing proteins, for further testing in development of new vaccine formulations to eliminate diseases caused by various pathogenic organisms. (M1) is useful for identifying putative adhesins that are important in drug discovery and preventing therapeutics for whooping cough, pneumonia, gastric ulcer and urinary tract infections. (M1) identifies adhesins from distantly related organisms, and from bacteria belonging to a wide phylogenetic spectrum. (M1) is capable of predicting adhesive nature of unique proteins. The present sequence is a microbial pathogen adhesin protein sequence.  
 CC  
 XX Sequence 431 AA;  
 SQ  
 Query Match 100.0%; Score 116; DB 9; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-10;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QYANYAGNQSIGNRGWFNP 20  
 Db |||||  
 403 QYANYAGNQSIGNRGWFNP 422  
 RESULT 10  
 ID ADD93651  
 XX ADD93651 standard; protein; 432 AA.  
 AC ADD93651;  
 XX 29-JAN-2004 (first entry)  
 DT  
 XX Streptococcus mutans glucan binding protein-B.  
 DE  
 XX Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.  
 KW Streptococcus mutans.  
 OS  
 XX WO2003075845-A2.  
 PN  
 XX 18-SEP-2003.  
 PD  
 XX 07-MAR-2003; 2003WO-US006962.  
 PF  
 XX 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 PR  
 XX (FORS-) FORSYTH INST.  
 PA Smith DJ, Taubman MA;  
 PI WPI; 2003-845091/78.  
 XX GENBANK; AY046412.  
 DR



OY 1 QEANYAGNOSIGNYRGWFNP 20  
 Db 404 QEANYAGNOSIGNYRGWFNP 423

## RESULT 13

ADX37275  
 ID ADX37275 standard; protein; 432 AA.

XX AC ADX37275;  
 XX 21-APR-2005 (first entry)  
 XX Streptococcus mutant glucan binding protein B variant #4.  
 DE immunogenicity; immune stimulation; glucan binding protein-B;  
 KW microparticle; major histocompatibility complex; tooth disease.  
 OS Streptococcus mutans.

XX US2005031633-A1.

PN 10-FEB-2005.

PD 09-MAR-2004; 2004US-00797821.

XX 13-APR-1998; 98US-0081550P.

PR 08-JAN-1999; 99US-0115142P.

PR 12-APR-1999; 99US-00290049.

PR 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

PR 07-MAR-2003; 2003US-00383930.

XX (SMIT/) SMITH D J.

PA (TAUB/) TAUBMAN M A.

PI Smith DJ, Taubman MA;

XX WPI; 2005-151644/16.

DR New composition comprising a fragment of a glucan binding protein-B  
 PT (GbpB) that binds to MHC class II protein, and a biocompatible  
 PT microparticle, useful for producing an antibody (claimed) for immunizing  
 PT mammals against dental caries.

XX Claim 3; SEQ ID NO 32; 73pp; English.

XX The invention relates to a composition comprising a fragment of a glucan  
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
 CC fragment binds to a major histocompatibility complex (MHC) class II  
 CC protein. The composition is useful for producing an antibody for  
 CC immunizing mammals against dental caries. This sequence corresponds to a  
 CC Streptococcus mutans GbpB protein of the invention.

XX Sequence 432 AA;

Query Match 100.0%; Score 116; DB 9; Length 432;

Best Local Similarity 100.0%; Pred. No. 8.6e-10;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QEANYAGNOSIGNYRGWFNP 20

Db 404 QEANYAGNOSIGNYRGWFNP 423

## RESULT 14

AAW60944

ID AAW60944 standard; protein; 132 AA.

XX AAW60944;

XX 13-OCT-1998 (first entry)

XX Streptococcus pneumoniae; type 4 strain.

XX WO200277021-A2.

DE Streptococcus pneumoniae encoded polypeptide.

XX coding region; ORF; open reading frame; antibacterial; infection;  
 KW prevention; meningitis.

XX Streptococcus pneumoniae.

XX WO9819689-A1.

XX 14-MAY-1998.

XX 27-OCT-1997; 97WO-US019226.

XX 01-NOV-1996; 96US-0029930P.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

PA (SMIK ) SMITHKLINE BEECHAM PLC.

XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;  
 PI Reid RH, Zarfos PN;

XX WPI; 1998-286586/25.

XX N-PSDB; AAV37344.

XX New isolated nucleic acids from Streptococcus pneumoniae - useful, e.g.  
 PT for identifying anti-bacterial(s) for treatment and prevention of  
 PT meningitis.

XX Claim 11; Page 47; 130pp; English.

XX The sequence is that of the polypeptide encoded by a region isolated from  
 CC S. pneumoniae. The protein, or agonists of it, may be useful as an  
 CC antibacterial for treatment or prevention of infection, specifically  
 CC caused by S. pneumoniae (particularly meningitis) but possibly also  
 CC Helicobacter pylori (ulcers and gastric cancer). It may be of particular  
 CC use before insertion of an in-dwelling device or any other invasive  
 CC procedure. The protein, or nucleic acid encoding it, can also be used in  
 CC vaccines to induce a cellular and/or humoral immune response, or to  
 CC screen for other antibacterials. The DNA may also contain flanking  
 CC sequences that are potential sources of control elements for bacterial  
 CC gene expression. Detecting a sequence encoding the protein can be used  
 CC diagnostically, e.g. to detect a mutation for serotyping or classifying  
 CC infectious agents

XX Sequence 132 AA;

Query Match 82.8%; Score 96; DB 2; Length 132;

Best Local Similarity 78.9%; Pred. No. 3.5e-07;

Matches 15; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 EANYAGNOSIGNYRGWFNP 20

Db 100 ESNYAGNRTIGNHRGWFP 118

## RESULT 15

ABU02747

ID ABU02747 standard; protein; 392 AA.

XX ABU02747;

XX 23-OCT-2003 (revised)

DT 11-FEB-2003 (first entry)

XX S. pneumoniae type 4 strain protein from coding region #2326.

XX Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
 KW gene therapy; vaccine.

XX Streptococcus pneumoniae; type 4 strain.

XX WO200277021-A2.

XX 03-OCT-2002.  
PD  
XX  
XX 27-MAR-2002; 2002WO-IB002163.  
PF  
XX  
XX 27-MAR-2001; 2001GB-00007658.  
PR  
XX  
XX (CHIR-) CHIRON SPA.  
PA  
XX (GENO-) INST GENOMIC RES.  
XX  
XX Masignani V, Tettelin H, Fraser C;  
PI  
XX  
XX WPI; 2003-040579/03.  
DR  
XX N-PSDB; ABX08038.  
DR  
XX  
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
PT ear infection.  
PT  
XX  
XX Claim 1; SEQ ID NO 4652; 56pp; English.  
PS  
XX  
XX The invention relates to a protein comprising or having at least 50%  
CC identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC ABS56454. Also included are an antibody which binds one of the proteins,  
CC treating a patient by administering the protein, DNA or antibody (in a  
CC composition), a kit comprising first and second primers, which are the  
CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
CC sequence not defined in the specification, for amplifying a target  
CC sequence contained within a Streptococcus nucleic acid sequence, where  
CC the first primer is substantially complementary to the target sequence  
CC and the second primer is substantially complementary to the complement of  
CC the target sequence, and where the parts of the primers having  
CC substantial complementarity define the termini of the target sequence to  
CC be amplified, assay comprising contacting a test compound with the  
CC protein, and determining whether the test compound binds to the protein  
CC and a Streptococcus pneumoniae bacterium, where one or more genes  
CC encoding the proteins has been rendered inactive. The proteins, nucleic  
CC acid molecules, antibody and compositions are useful as medicaments for  
CC treating or preventing a disease or infection due to streptococcus  
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
CC media or ear infection. They are also useful in developing vaccines,  
CC diagnostics and antibiotics. The methods are useful for identifying  
CC immunodominant proteins. The present sequence is one of the 2469 proteins  
CC expressed by the identified coding regions from the genomic sequence.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
CC standardise OS field)  
XX  
XX Sequence 392 AA;  
Query Match 82.8%; Score 96; DB 6; Length 392;  
Best Local Similarity 78.9%; Pred. No. 1.3e-06;  
Matches 15; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 2 EANYAGNQSIGNYRGWFNP 20  
|:|||||:|||||  
Db 360 ESNYAGNRTIGNHRCWFNP 378  
Search completed: February 15, 2006, 22:20:44  
Job time : 185 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 22:20:59 ; Search time 38 Seconds  
(without alignments)  
50.640 Million cell updates/sec

Title: US-10-797-821-22

Perfect score: 116

Sequence: 1 QEANYAGNQSIGNYRGWFNP 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 segs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 96    | 82.8        | 392    | 2 G95258 | secreted 45 kd pro |
| 2          | 96    | 82.8        | 392    | 2 B98124 | general stress pro |
| 3          | 55    | 47.4        | 332    | 2 AC0260 | hypothetical phage |
| 4          | 50    | 43.1        | 1061   | 2 A40609 | OmpA-related prote |
| 5          | 47    | 40.5        | 261    | 2 S71364 | ubiquinol-cytochro |
| 6          | 47    | 40.5        | 837    | 1 S54429 | outer membrane ugh |
| 7          | 46    | 39.7        | 319    | 1 S65523 | dihydrodipicolinat |
| 8          | 46    | 39.7        | 510    | 2 A98154 | hypothetical prote |
| 9          | 46    | 39.7        | 521    | 2 AC3134 | hypothetical prote |
| 10         | 46    | 39.7        | 619    | 1 KSNCL0 | laccase (EC 1.10.3 |
| 11         | 46    | 39.7        | 619    | 1 KSNCLT | laccase (EC 1.10.3 |
| 12         | 46    | 39.7        | 888    | 2 C90595 | hypothetical prote |
| 13         | 45.5  | 39.2        | 350    | 2 AG3425 | iolE protein limpo |
| 14         | 45.5  | 39.2        | 350    | 2 AE2954 | conserved hypothet |
| 15         | 45.5  | 39.2        | 370    | 2 H98328 | hypothetical prote |
| 16         | 45.5  | 39.2        | 385    | 2 D41732 | heterogeneous nucl |
| 17         | 45    | 38.8        | 108    | 1 KVM573 | Ig kappa chain v r |
| 18         | 45    | 38.8        | 242    | 1 F75433 | probable phosphoes |
| 19         | 45    | 38.8        | 514    | 2 B64634 | hypothetical prote |
| 20         | 45    | 38.8        | 514    | 2 E71880 | probable outer mem |
| 21         | 45    | 38.8        | 536    | 2 T35109 | hypothetical prote |
| 22         | 45    | 38.8        | 621    | 2 S72493 | laccase (EC 1.10.3 |
| 23         | 45    | 38.8        | 714    | 2 A83314 | probable chemotaxi |
| 24         | 45    | 38.8        | 835    | 2 T06590 | probable beta-gala |
| 25         | 44.5  | 38.4        | 1059   | 2 E87058 | isoleucyl-tRNA syn |
| 26         | 44    | 37.9        | 43     | 2 S21065 | Ig kappa chain v r |
| 27         | 44    | 37.9        | 147    | 2 I40625 | hypothetical prote |
| 28         | 44    | 37.9        | 237    | 2 F97109 | uncharacterized pr |
| 29         | 44    | 37.9        | 416    | 2 S09334 | transcobalamin I p |

|    |      |      |      |          |                    |
|----|------|------|------|----------|--------------------|
| 30 | 44   | 37.9 | 645  | 2 T16078 | hypothetical prote |
| 31 | 44   | 37.9 | 1006 | 2 S20126 | exoribonuclease RA |
| 32 | 43.5 | 37.5 | 160  | 2 JC7622 | actinohivin precu  |
| 33 | 43.5 | 37.5 | 430  | 2 S55325 | endo-beta-1,6-gluc |
| 34 | 43   | 37.1 | 125  | 2 S04935 | Ig kappa chain pre |
| 35 | 43   | 37.1 | 130  | 2 PLO113 | Ig kappa chain pre |
| 36 | 43   | 37.1 | 137  | 2 S50009 | hypothetical prote |
| 37 | 43   | 37.1 | 177  | 2 H81743 | conserved hypothet |
| 38 | 43   | 37.1 | 340  | 2 T26526 | hypothetical prote |
| 39 | 43   | 37.1 | 372  | 2 AG2827 | HFLK protein limpo |
| 40 | 43   | 37.1 | 373  | 2 E97605 | proteinase chain h |
| 41 | 43   | 37.1 | 451  | 1 S27617 | glucarate dehydrat |
| 42 | 43   | 37.1 | 516  | 2 T15633 | hypothetical prote |
| 43 | 43   | 37.1 | 548  | 2 T27542 | hypothetical prote |
| 44 | 43   | 37.1 | 614  | 2 S78277 | dnak-type molecula |
| 45 | 43   | 37.1 | 626  | 2 E84432 | hypothetical prote |

ALIGNMENTS

RESULT 1

G95258

secreted 45 kd protein [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004

C:Accession: G95258

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide  
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I.  
non, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: G95258

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-392 <KUR>

A:Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMY4; UNIPARC:UPI0000051B81; GB:AE005672; I

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP2216

Query Match 82.8%; Score 96; DB 2; Length 392;

Best Local Similarity 78.9%; Pred. No. 6.3e-07;

Matches 15; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 EANYAGNQSIGNYRGWFNP 20

I:|||||:|||||

DB 360 ESNYAGNRTIGNHRCWFNP 378

RESULT 2

B98124

General stress protein Gsp-781 [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004

C:Accession: B98124

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Es  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: B98124

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-392 <KUR>

A:Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMY4; UNIPARC:UPI0000051B81; GB:AE007317; I

C:Genetics:

A:Gene: gsp-781

Query Match 82.8%; Score 96; DB 2; Length 392;  
 Best Local Similarity 78.9%; Pred. No. 6.3e-07;  
 Matches 15; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFP 20  
 :|||||:|||||:  
 Db 360 ESNYAGNTIGNHRCWFP 378  
 :|||||:|||||:

RESULT 3  
 AC0260  
 hypochetical phage protein YPO2132 [imported] - Yersinia pestis (strain CO92)  
 C:Species: Yersinia pestis  
 C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
 C:Accession: AC0260  
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davis, P.; Dougan, G.;  
 il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,  
 Nature 413, 523-527, 2001  
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A:Reference number: AB0001; MUID:21470413; PMID:11586360  
 A:Accession: AC0260  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-332 <UNP>  
 A:Cross-references: UNIPROT:Q8ZEM9; UNIPARC:UPI00000CD8A6; GB:AL590842; PIDN:CAC90943.1;  
 C:Genetics:  
 A:Gene: YPO2132

Query Match 47.4%; Score 55; DB 2; Length 332;  
 Best Local Similarity 69.2%; Pred. No. 0.94;  
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 6 AGNQSIGNYRGWF 18  
 :|||:|||||:  
 Db 275 AGNINNGYRGWY 287  
 :|||:|||||:

RESULT 4  
 A40609  
 OmpA-related protein Oar - Myxococcus xanthus  
 C:Species: Myxococcus xanthus  
 C>Date: 16-Feb-1994 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
 C:Accession: A40609  
 R:Martinez-Canamero, M.; Munoz-Dorado, J.; Farez-Vidal, E.; Inouye, M.; Inouye, S.  
 J. Bacteriol. 175, 4756-4763, 1993  
 A:Title: Oar, a 115-kilodalton membrane protein required for development of Myxococcus x  
 A:Reference number: A40609; MUID:93328680; PMID:8335633  
 A:Accession: A40609  
 A:Status: preliminary  
 A:Molecule type: DNA; protein  
 A:Residues: 1-1061 <MAR>  
 A:Cross-references: UNIPROT:P38370; UNIPARC:UPI0000130B68; GB:S64103; NID:g402803; PIDN:  
 A:Note: sequence extracted from NCBI backbone (NCBIN:135644, NCBI:P:135645)

Query Match 43.1%; Score 50; DB 2; Length 1061;  
 Best Local Similarity 47.4%; Pred. No. 18;  
 Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFP 20  
 :|||:|||||:  
 Db 844 QANYTWSRLYGNYPGLFRP 862  
 :|||:|||||:

RESULT 5  
 S71364  
 ubiquinol-cytochrome-c reductase (EC 1.10.2.2) Rieske iron-sulfur protein precursor - Ch  
 N:Alternate names: Rieske iron-sulfur protein  
 C:Species: Chlamydomonas reinhardtii  
 C>Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 09-Jul-2004  
 C:Accession: S71364; S78209  
 R:Atteia, A.; Franzen, L.G.  
 Eur. J. Biochem. 237, 792-799, 1996

A:Title: Identification, cDNA sequence and deduced amino acid sequence of the mitochondri  
 g and subunit interaction.  
 A:Reference number: S71364; MUID:96235247; PMID:8647127  
 A:Accession: S71364  
 A:Molecule type: mRNA  
 A:Residues: 1-261 <ATT>  
 A:Cross-references: UNIPROT:Q42703; UNIPARC:UPI000009F24B; EMBL:X91795; NID:g1418673; PII  
 A:Experimental source: strain CW15  
 A:Accession: S78209  
 A:Molecule type: protein  
 A:Residues: 55-74 <ATY>  
 A:Cross-references: UNIPARC:UPI0000174C82  
 C:Genetics:  
 A:Genome: nuclear  
 C:Superfamily: ubiquinol-cytochrome-c reductase (cytochrome c1) (complex III), iron-sulfi  
 C:Keywords: 2Fe-2S; electron transfer; membrane-associated complex; metalloprotein; mito  
 F:1-54/Domain: transit peptide (mitochondrion) #status predicted <TNP>  
 F:55-261/Product: ubiquinol-cytochrome-c reductase iron-sulfur protein #status experiment  
 F:194-241/Domain: Rieske [2Fe-2S] homology <RSK>  
 F:204,206,223,226/Binding site: 2Fe-2S cluster (Cys, His, Cys, His) (covalent) #status p  
 F:209-225/Disulfide bonds: #status predicted  
 F:226/Active site: His #status predicted

Query Match 40.5%; Score 47; DB 2; Length 261;  
 Best Local Similarity 77.8%; Pred. No. 12;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 12 GNYRGWFP 20  
 :|||:|||||:  
 Db 216 GNYQGWFCP 224  
 :|||:|||||:

RESULT 6  
 S54429  
 outer membrane usher protein hifc precursor - Haemophilus influenzae (strain AM30)  
 C:Species: Haemophilus influenzae  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
 C:Accession: S54429  
 R:van Ham, S.M.; van Alphen, L.; Mooi, F.R.; van Putten, J.P.M.  
 Mol. Microbiol. 13, 673-684, 1994  
 A:Title: The fimbrial gene cluster of Haemophilus influenzae type b.  
 A:Reference number: S54428; MUID:95089703; PMID:7997179  
 A:Accession: S54429  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-837 <VAN>  
 A:Cross-references: UNIPROT:P45997; UNIPARC:UPI000012C67D; EMBL:Z33502; NID:G535165; PID  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1994  
 C:Genetics:  
 A:Gene: hifc  
 A:Start codon: GTG  
 C:Superfamily: outer membrane usher protein fimD  
 C:Keywords: membrane protein

Query Match 40.5%; Score 47; DB 1; Length 837;  
 Best Local Similarity 50.0%; Pred. No. 41;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFP 19  
 :|||:|||||:  
 Db 623 DANLAHNSIGSYRASYS 640  
 :|||:|||||:

RESULT 7  
 S56523  
 dihydrodipicolinate synthase homolog yjH - Escherichia coli (strain K-12)  
 C:Species: Escherichia coli  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 01-Mar-2002  
 C:Accession: S56523; D65243  
 R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
 Nucleic Acids Res. 23, 2105-2119, 1995  
 A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.1  
 A:Reference number: S56314; MUID:95334362; PMID:7610040

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AC3134

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-521 <KUR>

A:Cross-references: UNIPROT:Q8UGW1; UNIPARC:UPI00000D267E; GB:AE008689; PIDN:AAL45489.1;

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu4695

A:Map position: linear chromosome

Query Match 39.7%; Score 46; DB 2; Length 521;

Best Local Similarity 50.0%; Pred. No. 35;

Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 7 GNQSIGNRGWFNP 20

Db 437 GAMGVSYGGWSNP 450

RESULT 10

KSNCL0

laccase (EC 1.10.3.2) precursor - Neurospora crassa (strain OR)

N/Alternate names: urishiol oxidase

C/Species: Neurospora crassa

C/Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004

C/Accession: A28523; A29762

R/Germann, U.A.; Mueller, G.; Hunziker, P.E.; Lersch, K.

J. Biol. Chem. 263, 885-896; 1988

A/Title: Characterization of two allelic forms of Neurospora crassa laccase. Amino- and

A/Reference number: A28523; MUID:88087214; PMID:2961749

A:Accession: A28523

A:Molecule type: DNA

A:Residues: 1-619 <GER>

A:Cross-references: UNIPROT:P06811; UNIPARC:UPI00001721B1; EMBL:M14554

R/Germann, U.A.; Lersch, K.

Proc. Natl. Acad. Sci. U.S.A. 83, 8854-8858, 1986

A/Title: Isolation and partial nucleotide sequence of the laccase gene from Neurospora

A/Reference number: A29762; MUID:87067412; PMID:2947240

A:Accession: A29762

A:Molecule type: DNA

A:Residues: 379-619 <GE2>

A:Cross-references: UNIPARC:UPI0000168938; GB:M14554; NTD:G168823; PIDN:AAA33590.1; PID:Q

C/Comment: This enzyme, which catalyzes the oxidation of benzidiol to benzosemiquinone

C:Genetics:

A:Introns: 86/3

C:Superfamily: laccase

C/Keywords: copper; glycoprotein; oxidoreductase

F/1-21/Domain: signal sequence #status predicted <SIG>

F/22-49/Domain: propeptide #status predicted <PRO>

F/50-619/Product: laccase #status predicted <PRO>

F/84-215/Domain: amino-terminal beta-barrel #status predicted <NAT>

F/216-372/Domain: middle beta-barrel #status predicted <BB1>

F/431-580/Domain: carboxyl-terminal beta-barrel #status predicted <BB2>

F/139,282,295,340,422,444/Binding site: carbohydrate (Asn) (covalent) #status predicted

F/144,480/Binding site: copper (His) (type 2) #status predicted

F/146,489,491,482,548,550/Binding site: 2Cu-O cluster (His) (type 3) #status predicted

F/477,549,554/Binding site: copper (His, Cys, His) (type 1) #status predicted

Query Match 39.7%; Score 46; DB 1; Length 619;

Best Local Similarity 66.7%; Pred. No. 42;

Matches 8; Conservative 2; Mismatches 0; Indels 2; Gaps 1;

Qy 8 NQSIGNRGWFNP 19

Db 331 NQAVGNY--WFN 340

RESULT 11

KSNCLT

laccase (EC 1.10.3.2) precursor - Neurospora crassa (strain TS)

N/Alternate names: urishiol oxidase

C/Species: Neurospora crassa

C>Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
 C/Accession: B28523  
 J/Germann, U.A.; Mueller, G.; Hunziker, P.E.; Lerch, K.  
 R. Biol. Chem. 263, 885-896, 1988  
 A/Title: Characterization of two allelic forms of *Neurospora crassa* laccase. Amino- and  
 A/Reference number: A28523; MUID:88087314; PMID:2961749  
 A/Accession: B28523  
 A/Molecule type: DNA  
 A/Residues: 1-619 <GER>  
 A/Cross-references: UNIPROT:P10574; UNIPARC:UPI000012E1B2; EMBL:M18334; NID:g168827; PID  
 C/Comment: This enzyme, which catalyzes the oxidation of benzendiol to benzosemiquinone  
 C/Genetics:  
 A/Introns: 86/3  
 C/Superfamily: laccase  
 C/Keywords: copper; glycoprotein; oxidoreductase  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-49/Domain: propeptide #status predicted <PRO>  
 F:50-619/Product: laccase #status predicted <MAT>  
 F:84-215/Domain: amino-terminal beta-barrel #status predicted <BB1>  
 F:216-372/Domain: middle beta-barrel #status predicted <BB2>  
 F:431-580/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>  
 F:139,282,295,340,422,444/Binding site: carbohydrate (Aen) (covalent) #status predicted  
 F:144,480/Binding site: copper (His) (type 2) #status predicted  
 F:146,189,191,482,548,550/Binding site: 2Cu-O cluster (His) (copper type 3) #status pred  
 F:477,549,554/Binding site: copper (His, Cys, His) (type 1) #status predicted

Query Match 39.7%; Score 46; DB 1; Length 619;  
 Best Local Similarity 66.7%; Pred. No. 42;  
 Matches 8; Conservative 2; Mismatches 0; Indels 2; Gaps 1;

QY 8 NOSIGNYRGWEN 19  
 ||:|||||  
 DB 331 NOAVGNV--WFN 340

RESULT 12  
 C90595  
 hypothetical protein MYPV\_6670 [imported] - Mycoplasma pulmonis (strain UAB CTIP)  
 C/Species: Mycoplasma pulmonis  
 C/Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
 R/Accession: C90595  
 C/Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;  
 Nucleic Acids Res. 29, 2145-2153, 2001  
 A/Title: The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulm*  
 A/Reference number: A99512; MUID:21267165; PMID:1133084  
 A/Accession: C90595  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-888 <KUR>  
 A/Cross-references: UNIPROT:Q98PQ2; UNIPARC:UPI00000D463E; GB:AL445566; PID:g14090082; F  
 A/Experimental source: strain UAB CTIP  
 C/Genetics:  
 A/Gene: MYPV\_6670  
 A/Genetic code: SGC3  
 C/Superfamily: isoleucine-tRNA ligase

Query Match 39.7%; Score 46; DB 2; Length 888;  
 Best Local Similarity 63.6%; Pred. No. 61;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 9 QSIGNYRGWEN 19  
 :|||||  
 DB 551 EGIDQYRGWEN 561

RESULT 13  
 AG3425  
 iolE protein [imported] - *Brucella melitensis* (strain 16M)  
 C/Species: *Brucella melitensis*  
 C/Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 09-Jul-2004  
 C/Accession: AG3425  
 R/DeiVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
 ; Mazur, M.; Goltzman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letess

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A/Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
 A/Reference number: AD3252; PMID:11756688  
 A/Accession: AG3425  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-350 <KUR>  
 A/Cross-references: UNIPROT:Q8VFX3; UNIPARC:UPI0000058056; GB:AE008917; PIDN:AAL52570.1;  
 A/Experimental source: strain 16M  
 C/Genetics:  
 A/Gene: BMEI1389  
 A/Map position: I

Query Match 39.2%; Score 45.5; DB 2; Length 350;  
 Best Local Similarity 47.4%; Pred. No. 28;  
 Matches 9; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 2 EANYAGNQSI-GNYRGWEN 19  
 |||||:|||||  
 DB 248 EFNPTGRQGVVGGYQGWV 266

RESULT 14  
 AB2954  
 conserved hypothetical protein Atu3235 [imported] - *Agrobacterium tumefaciens* (strain C58  
 C/Species: *Agrobacterium tumefaciens*  
 C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 09-Jul-2004  
 C/Accession: AB2954  
 R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, I.  
 ster, E.W.  
 A/Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
 A/Reference number: AB2577; MUID:21608550; PMID:11743193  
 A/Accession: AB2954  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-350 <KUR>  
 A/Cross-references: UNIPROT:Q8UAY2; UNIPARC:UPI00001647C7; GB:AE008689; PIDN:AAL44051.1;  
 A/Experimental source: strain C58 (Dupont)  
 C/Genetics:  
 A/Gene: Atu3235  
 A/Map position: linear chromosome

Query Match 39.2%; Score 45.5; DB 2; Length 350;  
 Best Local Similarity 47.4%; Pred. No. 28;  
 Matches 9; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 2 EANYAGNQSI-GNYRGWEN 19  
 |||||:|||||  
 DB 248 EFNPTGRQGVVGGYQGWV 266

RESULT 15  
 H98328  
 hypothetical protein AGR\_L\_3161 [imported] - *Agrobacterium tumefaciens* (strain C58, Cerec  
 C/Species: *Agrobacterium tumefaciens*  
 C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
 C/Accession: H98328  
 R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman,  
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
 Science 294, 2323-2328, 2001  
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tum*  
 A/Reference number: A97359; MUID:21608551; PMID:11743194  
 A/Accession: H98328  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-370 <KUR>  
 A/Cross-references: UNIPROT:Q8UAY2; UNIPARC:UPI00000D211E; GB:AE007870; PIDN:AAK90154.1;  
 C/Genetics:  
 A/Gene: AGR\_L\_3161



A:Map position: linear chromosome

Query Match 39.2%; Score 45.5; DB 2; Length 370;  
Best Local Similarity 47.4%; Pred. No. 29;  
Matches 9; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 2 EANYAGNOSI-GNYRGWFN 19  
| | | | | : | | | |  
Db 268 EFNPTGRQGVGGYQGVN 286

Search completed: February 15, 2006, 22:25:22  
Job time : 40 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 22:17:44 ; Search time 228 Seconds  
(without alignments)  
61.888 Million cell updates/sec

Title: US-10-797-821-22

Perfect score: 116

Sequence: 1 Q9ANYAGNQSIGNVRGWFNP 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_05.80.\*

1: uniprot\_eprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description  |
|------------|-------|-------------|--------|-------|--------------|
| 1          | 116   | 100.0       | 431    | 2     | Q938V0 STRMU |
| 2          | 116   | 100.0       | 431    | 2     | Q938V3 STRMU |
| 3          | 116   | 100.0       | 431    | 2     | Q9AG98 STRMU |
| 4          | 116   | 100.0       | 431    | 2     | Q8DMW3 STRMU |
| 5          | 116   | 100.0       | 432    | 2     | Q938V1 STRMU |
| 6          | 116   | 100.0       | 432    | 2     | Q938V2 STRMU |
| 7          | 96    | 82.8        | 392    | 2     | Q8DMY4 STRR6 |
| 8          | 96    | 82.8        | 392    | 2     | Q97N55 STRPN |
| 9          | 89    | 76.7        | 447    | 2     | Q9AKA4 STRAG |
| 10         | 89    | 76.7        | 447    | 2     | Q8E2H1 STRA5 |
| 11         | 89    | 76.7        | 447    | 2     | Q8E7X9 STRA3 |
| 12         | 84    | 72.4        | 169    | 2     | Q8DXT4 STRA5 |
| 13         | 84    | 72.4        | 169    | 2     | Q8E3F4 STRA3 |
| 14         | 79    | 68.1        | 129    | 2     | Q5M137 STRT1 |
| 15         | 79    | 68.1        | 461    | 2     | Q56SA7 STRTR |
| 16         | 79    | 68.1        | 474    | 2     | Q5M6K4 STRT2 |
| 17         | 79    | 68.1        | 482    | 2     | Q5M5M6 STRT2 |
| 18         | 79    | 68.1        | 485    | 2     | Q5W213 STRT1 |
| 19         | 77    | 66.4        | 398    | 2     | Q5XEL1 STRP6 |
| 20         | 77    | 66.4        | 398    | 2     | Q9A1Z8 STRPY |
| 21         | 77    | 66.4        | 398    | 2     | Q7CNQ7 STRP8 |
| 22         | 77    | 66.4        | 398    | 2     | Q8P318 STRP3 |
| 23         | 74    | 63.8        | 544    | 2     | Q840V8 STRMU |
| 24         | 74    | 63.8        | 544    | 2     | Q840W6 STRMU |
| 25         | 74    | 63.8        | 544    | 2     | Q840X3 STRMU |
| 26         | 74    | 63.8        | 544    | 2     | Q8DUR7 STRMU |
| 27         | 70    | 60.3        | 211    | 2     | Q8DVU8 STRMU |
| 28         | 70    | 60.3        | 226    | 2     | Q93RG6 STRIT |
| 29         | 55    | 47.4        | 332    | 2     | Q8ZEM9 YERPE |
| 30         | 55    | 47.4        | 337    | 2     | Q8CL50 YERPE |
| 31         | 53.5  | 46.1        | 906    | 2     | Q55KP5 CRYNE |

|    |      |      |      |   |              |                     |
|----|------|------|------|---|--------------|---------------------|
| 32 | 53.5 | 46.1 | 906  | 2 | QSKAT1 CRYNE | Qskat1 cryptococcus |
| 33 | 52   | 44.8 | 188  | 2 | Q82CS4 STRAW | Q82cs4 streptomyces |
| 34 | 50   | 43.1 | 1061 | 1 | OAR MYXA     | P38370 myxococcus   |
| 35 | 49.5 | 42.7 | 454  | 2 | O5CNL2 CRYHO | Q5cnl2 cryptospori  |
| 36 | 49.5 | 42.7 | 950  | 2 | O4XVZ4 PLACH | O4xvz4 plasmodium   |
| 37 | 49   | 42.2 | 267  | 2 | Q4HU65 GIBZE | Q4hu65 gibberella   |
| 38 | 49   | 42.2 | 523  | 2 | Q6L9J3 9GAMM | Q6l9j3 aeromonas v  |
| 39 | 49   | 42.2 | 588  | 2 | Q9RMM8 AERHY | Q9rmm8 aeromonas h  |
| 40 | 49   | 42.2 | 590  | 2 | Q9R9S7 AERPU | Q9r9s7 aeromonas p  |
| 41 | 49   | 42.2 | 595  | 2 | Q9LCJ5 AERPU | Q9lcj5 aeromonas p  |
| 42 | 48.5 | 41.8 | 189  | 2 | O6ZIC2 ORYSA | O6zic2 oryza sativ  |
| 43 | 48   | 41.4 | 385  | 2 | Q5DD85 SCHJA | Q5dd85 schistosoma  |
| 44 | 48   | 41.4 | 469  | 2 | Q5K163 GEOKA | Q5ky63 geobacillus  |
| 45 | 48   | 41.4 | 489  | 2 | Q8GCX0 CHRVO | Q8gxc0 chromobacte  |

ALIGNMENTS

RESULT 1  
Q938V0 STRMU  
ID Q938V0\_STRMU PRELIMINARY; PRT; 431 AA.  
AC Q938V0\_STRMU PRELIMINARY; PRT; 431 AA.  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE Glucan-binding protein B.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=5SM3;  
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;  
RT "Cloning of the gbpB gene from Streptococcus mutans.";  
RL J. Dent. Res. 79:224-224 (2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=5SM3;  
RX MEDLINE=21481971; PubMed=11598068;  
RX DOI=10.1128/IAI.69.11.6931-6941.2001;  
RA Mattos-Granier R.O., Jin S., King W.F., Chen T., Smith D.J.,  
Duncan M.J.;  
RT "Cloning of the Streptococcus mutans gene encoding glucan binding  
protein B and analysis of genetic diversity and protein production in  
clinical isolates.";  
RL Infect. Immun. 69:6931-6941 (2001).  
DR EMBL; AY046414; AAK94504.1; -; Genomic\_DNA.  
DR InterPro; IPR007921; CHAP.  
DR InterPro; IPR009148; Siba.  
DR Pfam; PF05257; CHAP; 1.  
DR PRINTS; PR01852; SIBAPROTEIN.  
DR PROSITE; PS50911; CHAP; 1.  
SQ SEQUENCE 431 AA; 44650 MW; 05D38D8B8C4609F CRC64;  
Query Match 100.0%; Score 116; DB 2; Length 431;  
Best Local Similarity 100.0%; Pred. No. 1.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 Q9ANYAGNQSIGNVRGWFNP 20  
|||||  
DB 403 Q9ANYAGNQSIGNVRGWFNP 422  
RESULT 2  
Q938V3 STRMU  
ID Q938V3\_STRMU PRELIMINARY; PRT; 431 AA.  
AC Q938V3;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE Glucan-binding protein B.

OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]\_ NUCLEOTIDE SEQUENCE.  
 RP STRAIN=3J32;  
 RC STRAIN=3J32;  
 RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;  
 RT "Cloning of the gbpB gene from Streptococcus mutans."  
 RL J. Dent. Res. 79:224-224 (2000).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=3J32;  
 RX MEDLINE=21481977; PubMed=11598068;  
 RA DOI=10.1128/IAI.69.11.6931-6941.2001;  
 RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,  
 RA Duncan M.J.;  
 RT "Cloning of the Streptococcus mutans gene encoding glucan binding  
 RT protein B and analysis of genetic diversity and protein production in  
 RT clinical isolates."  
 RL Infect. Immun. 69:6931-6941 (2001).  
 DR EMBL; AY046410; AAK94500.1; -; Genomic\_DNA.  
 DR InterPro; IPR007921; CHAP.  
 DR InterPro; IPR009148; Siba.  
 DR Pfam; PF05257; CHAP; 1.  
 DR PRINTS; PRO1852; SIBAPROTEIN.  
 DR PROSITE; PS50911; CHAP; 1.  
 SQ SEQUENCE 431 AA; 44620 MW; 464FE3B563FB7E51 CRC64;  
 Query Match 100.0%; Score 116; DB 2; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-09; Indels 0; Gaps 0;  
 Matches 20; Conservative 0; Mismatches 0;  
 QY 1 QYANYAGNQSIGNYRGWFPN 20  
 DB 403 QYANYAGNQSIGNYRGWFPN 422

RESULT 3  
 Q9AG98\_STRMU  
 ID Q9AG98\_STRMU PRELIMINARY; PRT; 431 AA.  
 AC Q9AG98;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)  
 DE Immunodominant glycoprotein IDG-60 (Glucan-binding protein B).  
 GN Name=saga;  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=GS-5;  
 RX MEDLINE=21153617; PubMed=11254612;  
 RA DOI=10.1128/IAI.69.4.2493-2501.2001;  
 RA Chia J.S., Lee Y.Y., Huang P.T., Chen J.Y.;  
 RT "Identification of stress-responsive genes in Streptococcus mutans by  
 RT differential display reverse transcription-PCR."  
 RL Infect. Immun. 69:2493-2501 (2001).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=GS-5;  
 RX MEDLINE=21481977; PubMed=11598074;  
 RA DOI=10.1128/IAI.69.11.6987-6998.2001;  
 RA Chia J.S., Chang L.Y., Shun C.T., Chang Y.Y., Chen J.Y.;  
 RT "A 60-kilodalton immunodominant glycoprotein is essential for cell  
 RT wall integrity and the maintenance of cell shape in Streptococcus  
 RT mutans."  
 RL Infect. Immun. 69:6987-6998 (2001).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=GS-5;

RA Chia J.-S., Chang L.-Y., Lee Y.-Y., Chen J.-Y.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=3VF4;  
 RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;  
 RT "Cloning of the gbpB gene from Streptococcus mutans."  
 RL J. Dent. Res. 79:224-224 (2000).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=3VF4;  
 RX MEDLINE=21481971; PubMed=11598068;  
 RA DOI=10.1128/IAI.69.11.6931-6941.2001;  
 RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,  
 RA Duncan M.J.;  
 RT "Cloning of the Streptococcus mutans gene encoding glucan binding  
 RT protein B and analysis of genetic diversity and protein production in  
 RT clinical isolates."  
 RL Infect. Immun. 69:6931-6941 (2001).  
 DR EMBL; AF338445; AAK08104.1; -; Genomic\_DNA.  
 DR EMBL; AY046411; AAK94501.1; -; Genomic\_DNA.  
 DR InterPro; IPR007921; CHAP.  
 DR InterPro; IPR009148; Siba.  
 DR Pfam; PF05257; CHAP; 1.  
 DR PRINTS; PRO1852; SIBAPROTEIN.  
 DR PROSITE; PS50911; CHAP; 1.  
 SQ SEQUENCE 431 AA; 44592 MW; 3BEB21FC5E47232E CRC64;  
 Query Match 100.0%; Score 116; DB 2; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-09; Indels 0; Gaps 0;  
 Matches 20; Conservative 0; Mismatches 0;  
 QY 1 QYANYAGNQSIGNYRGWFPN 20  
 DB 403 QYANYAGNQSIGNYRGWFPN 422

RESULT 4  
 Q8DWM3\_STRMU  
 ID Q8DWM3\_STRMU PRELIMINARY; PRT; 431 AA.  
 AC Q8DWM3;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Putative secreted antigen GbpB/SagA; putative peptidoglycan  
 DE hydrolase.  
 GN Name=gbpB; OrderedLocusNames=SMU.22;  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=UA159 / ATCC 700610 / Serotype c;  
 RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;  
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,  
 RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,  
 RA Ferretti J.J.;  
 RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental  
 RT pathogen."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).  
 DR EMBL; AE014855; AAN57811.1; -; Genomic\_DNA.  
 DR InterPro; IPR007921; CHAP.  
 DR InterPro; IPR009148; Siba.  
 DR Pfam; PF05257; CHAP; 1.  
 DR PRINTS; PRO1852; SIBAPROTEIN.  
 DR PROSITE; PS50911; CHAP; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 431 AA; 44620 MW; 2DICA685248CCD3E CRC64;  
 Query Match 100.0%; Score 116; DB 2; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-09; Indels 0; Gaps 0;

```

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QYANYAGNOSIGNYRGWFNP 20
|||||
DB 403 QYANYAGNOSIGNYRGWFNP 422

RESULT 5
Q938V1_STRMU
ID Q938V1_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V1_STRMU
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-NAR-2004 (TREMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RX MEDLINE=21481971; PubMed=11598068;
DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941 (2001).
DR EMBL; AY046413; AAK94503.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44652 MW; 3F88ECB9A1F3BE4F CRC64;

Query Match 100.0%; Score 116; DB 2; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QYANYAGNOSIGNYRGWFNP 20
|||||
DB 404 QYANYAGNOSIGNYRGWFNP 423

RESULT 6
Q938V2_STRMU
ID Q938V2_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V2_STRMU
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-NAR-2004 (TREMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=15JP2;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.

```

```

RC STRAIN=15JP2;
RX MEDLINE=21481971; PubMed=11598068;
DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941 (2001).
DR EMBL; AY046413; AAK94503.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44648 MW; E769B2504AE50E9 CRC64;

Query Match 100.0%; Score 116; DB 2; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QYANYAGNOSIGNYRGWFNP 20
|||||
DB 404 QYANYAGNOSIGNYRGWFNP 423

RESULT 7
Q8DMV4_STRR6
ID Q8DMV4_STRR6 PRELIMINARY; PRT; 392 AA.
AC Q8DMV4_STRR6
DT 01-NAR-2003 (TREMBLrel. 23, Created)
DT 01-NAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-NAR-2004 (TREMBLrel. 26, Last annotation update)
DE General stress protein GSP-781.
GN Name=gsp-781; OrderedLocusNames=spr2021;
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21429245; PubMed=11544234;
DOI=10.1128/JB.183.19.5709-5717.2001;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Bargett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren D.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717 (2001).
DR EMBL; AE008566; AAL00823.1; -; Genomic_DNA.
DR PIR; B98124; B98124.
DR PIR; G95258; G95258.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 392 AA; 41697 MW; B5A5A860FAEA77DD CRC64;

Query Match 82.8%; Score 96; DB 2; Length 392;
Best Local Similarity 78.9%; Pred. No. 2.1e-06;
Matches 15; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 2 BANYAGNOSIGNYRGWFNP 20
|||||
DB 360 BANYAGNOSIGNYRGWFNP 378

```

```
RESULT 8
Q97N55_STRPN PRELIMINARY; PRT; 392 AA.
AC Q97N55;
DT 01-OCT-2001 (TRENBLrel. 18, Created)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Secreted 45 kd protein.
OS Streptococcus pneumoniae.
NCBI_TaxID=1313;
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA Tetelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA Hickey E.K., Holt I.E., Loftus B.J., Yang P., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AE007509; AAK76264.1; -; Genomic_DNA.
DR PIR; B98124; B98124.
DR PIR; G95258; G95258.
DR TIGR; SP2216; -.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 392 AA; 41697 MW; B5A5A860FAEA77DD CRC64;

Query Match 82.8%; Score 96; DB 2; Length 392;
Best Local Similarity 78.9%; Pred. No. 2.1e-06;
Matches 15; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFNP 20
|:|||||:|||||
Db 360 ESNYAGNRTGNHRCWFNP 378

RESULT 9
Q9AKA4_STRAG PRELIMINARY; PRT; 447 AA.
AC Q9AKA4;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE PcsB protein precursor.
GN Name=pcsb;
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=1311;
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=6313;
RX MEDLINE=21101799; PubMed=11157929;
RX DOI=10.1128/JB.183.4.1175-1183.2001;
RA Reinscheid D.J., Gottschalk B., Schubert A., Eikmanns B.J.,
RA Chhatwal G.S.;
RT "Identification and molecular analysis of PcsB, a protein required for
RT cell wall separation of group B streptococcus."
RL J. Bacteriol. 183:1175-1183(2001).
```

```
DR EMBL; AJ277292; CAC28144.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Signal.
FT SIGNAL 1 25 Potential.
FT CHAIN 26 447 PcsB protein.
SQ SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 76.7%; Score 89; DB 2; Length 447;
Best Local Similarity 84.2%; Pred. No. 3.1e-05;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFNP 20
|:|||||:|||||
Db 417 ESNYAGNMSIGNYRGSFNP 435

RESULT 10
Q8E2H1_STRAS PRELIMINARY; PRT; 447 AA.
ID Q8E2H1_STRAS PRELIMINARY;
AC Q8E2H1;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE PcsB protein.
GN Name=pcsb; OrderedLocNames=SAG0017;
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=216466;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Tetelin H., Massignani V., Cieslewicz M.J., Eisen J.A., Peterson S.N.,
RA Wesels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V streptococcus agalactiae."
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014192; AAM98925.1; -; Genomic_DNA.
DR TIGR; SAG0017; -.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 76.7%; Score 89; DB 2; Length 447;
Best Local Similarity 84.2%; Pred. No. 3.1e-05;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFNP 20
|:|||||:|||||
Db 417 ESNYAGNMSIGNYRGSFNP 435

RESULT 11
Q8E7X9_STRAS3 PRELIMINARY; PRT; 447 AA.
ID Q8E7X9_STRAS3 PRELIMINARY;
AC Q8E7X9;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
```

```

DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE pcsB protein.
GN Name=pcsB; OrderedLocusNames=gb0016;
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rueniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513 (2002).
DR EMBL; AL766843; CAD45661.1; -; Genomic_DNA.
DR SagalList; gbs0016; -.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 76.7%; Score 89; DB 2; Length 447;
Best Local Similarity 84.2%; Pred. No. 3.1e-05;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFNP 20
DB 417 ESNYAGNWSIGNYRGSFNP 435

RESULT 12
QSDXT4 STRA5
ID Q8DXT4_STRA5 PRELIMINARY; PRT; 169 AA.
AC Q8DXT4;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Hypothetical protein SAG1762.
GN OrderedLocusNames=SAG1762;
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Wesselin H., Maignani V., Cieleswicz M.J., Eisen J.A., Peterson S.N.,
RA Tettelin H., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.S., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,
RA Carthy H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
DR EMBL; AE014271; RAN00625.1; -; Genomic_DNA.
DR TIGR; SAG1762; -.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.

KW Complete proteome.
SQ SEQUENCE 169 AA; 18390 MW; 8A1653A8B5B8B769 CRC64;

Query Match 72.4%; Score 84; DB 2; Length 169;
Best Local Similarity 73.7%; Pred. No. 6.4e-05;
Matches 14; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFNP 20
DB 136 ESNYAGHQWIDNYRGWFD 154

RESULT 13
Q8E3F4 STRA3
ID Q8E3F4_STRA3 PRELIMINARY; PRT; 169 AA.
AC Q8E3F4;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Hypothetical protein gbs1805.
GN OrderedLocusNames=gbs1805;
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rueniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513 (2002).
DR EMBL; AL766853; CAD47464.1; -; Genomic_DNA.
DR SagalList; gbs1805; -.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 169 AA; 18304 MW; 8A159F753D747869 CRC64;

Query Match 72.4%; Score 84; DB 2; Length 169;
Best Local Similarity 73.7%; Pred. No. 6.4e-05;
Matches 14; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFNP 20
DB 136 ESNYAGHQWIDNYRGWFD 154

RESULT 14
QSM137 STRT1
ID QSM137_STRT1 PRELIMINARY; PRT; 129 AA.
AC QSM137;
DT 01-FEB-2005 (TremBLrel. 29, Created)
DT 01-FEB-2005 (TremBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TremBLrel. 29, Last annotation update)
DE Glucan binding protein.
GN OrderedLocusNames=st0443;
OS Streptococcus thermophilus (strain CNRZ 1066).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=299768;
RN [1]_
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15543133; DOI=10.1038/nbt1034;
RA Bolotin A., Quinquis B., Renault P., Sorokin A., Ehrlich S.D.,
RA Kulakauskas S., Lapidus A., Goldsman E., Mazur M., Pusch G.D.,
RA Fonstein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D.,

```

RA Ngui K., Masuy D., Hancy F., Burtreau S., Boutry M., Delcours J.,  
 RA Goffeau A., Hols P.;  
 RT "Complete sequence and comparative genome analysis of the dairy  
 RT bacterium Streptococcus thermophilus.";  
 RL Nat. Biotechnol. 22:1554-1558(2004).  
 DR EMBL: CP000024; AAV62043.1; -; Genomic\_DNA.  
 DR InterPro: IPR007921; CHAP.  
 DR InterPro: IPR009148; SIDA.  
 DR Pfam: PF05257; CHAP; 1.  
 DR PRINTS: PRO1852; SIAPROTEIN.  
 DR PROSITE: PS50911; CHAP; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 129 AA; 13136 MW; 7ADB47F41B7D4B3 CRC64;

Query Match 68.1%; Score 79; DB 2; Length 129;  
 Best Local Similarity 73.7%; Pred. No. 0.0029;  
 Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFNP 20  
 Db 97 ESNYAGNMLIGNYRGTFDP 115

RESULT 15

ID Q56SA7\_STRTR PRELIMINARY; PRT; 461 AA.  
 AC Q56SA7;  
 DT 10-MAY-2005 (TREMBlrel. 30, Created)  
 DT 10-MAY-2005 (TREMBlrel. 30, Last sequence update)  
 DT 10-MAY-2005 (TREMBlrel. 30, Last annotation update)  
 DE Cse.  
 OS Streptococcus thermophilus.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1308;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=CNRZ368;  
 RA Borges F., Layec S., Thibessard A., Fernandez A., Gintz B., Hols P.,  
 RA Decaris B., Leblond-Bourget N.;  
 RT "cse, a Chimeric and Variable Gene, Encodes an Extracellular Protein  
 RT Involved in Cellular Segregation in Streptococcus thermophilus.";  
 RL J. Bacteriol. 187:2737-2746(2005).  
 DR EMBL: AY695844; AAW33692.1; -; Genomic DNA.  
 SQ SEQUENCE 461 AA; 48498 MW; CA72A95F945BD7A4 CRC64;

Query Match 68.1%; Score 79; DB 2; Length 461;  
 Best Local Similarity 73.7%; Pred. No. 0.0012;  
 Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFNP 20  
 Db 429 ESNYAGNMLIGNYRGTFDP 447

Search completed: February 15, 2006, 22:24:38  
 Job time : 230 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 22:36:04 ; Search time 165 Seconds  
(without alignments)  
50.646 Million cell updates/sec

Title: US-10-797-821-22

Perfect score: 116

Sequence: 1 QEANYAGNQSIGNYRGWFNP 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA\_Main:

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description          |
|------------|-------|-------------|--------|-------|----------------------|
| 1          | 116   | 100.0       | 20     | 4     | US-10-383-930-22     |
| 2          | 116   | 100.0       | 20     | 5     | US-10-797-821-22     |
| 3          | 116   | 100.0       | 431    | 4     | US-10-383-930-29     |
| 4          | 116   | 100.0       | 431    | 4     | US-10-383-930-30     |
| 5          | 116   | 100.0       | 431    | 4     | US-10-383-930-33     |
| 6          | 116   | 100.0       | 431    | 5     | US-10-797-821-29     |
| 7          | 116   | 100.0       | 431    | 5     | US-10-797-821-30     |
| 8          | 116   | 100.0       | 431    | 5     | US-10-797-821-33     |
| 9          | 116   | 100.0       | 432    | 4     | US-10-383-930-31     |
| 10         | 116   | 100.0       | 432    | 4     | US-10-383-930-32     |
| 11         | 116   | 100.0       | 432    | 5     | US-10-797-821-31     |
| 12         | 116   | 100.0       | 432    | 5     | US-10-797-821-32     |
| 13         | 96    | 82.8        | 392    | 5     | US-10-472-928-4652   |
| 14         | 96    | 82.8        | 399    | 5     | US-10-617-320-3230   |
| 15         | 77    | 66.4        | 398    | 5     | US-10-474-792-600    |
| 16         | 52    | 44.8        | 188    | 4     | US-10-156-761-12798  |
| 17         | 49    | 42.2        | 618    | 4     | US-10-425-115-325660 |
| 18         | 47    | 40.5        | 198    | 4     | US-10-425-115-304530 |
| 19         | 47    | 40.5        | 290    | 4     | US-10-424-599-196081 |
| 20         | 47    | 40.5        | 509    | 4     | US-10-437-963-105017 |
| 21         | 47    | 40.5        | 512    | 4     | US-10-425-115-302311 |
| 22         | 47    | 40.5        | 524    | 4     | US-10-425-114-68132  |
| 23         | 47    | 40.5        | 838    | 5     | US-10-836-164-10     |
| 24         | 46    | 39.7        | 25     | 3     | US-09-847-102A-107   |
| 25         | 46    | 39.7        | 25     | 3     | US-09-847-102A-108   |
| 26         | 46    | 39.7        | 123    | 4     | US-10-425-115-347949 |
| 27         | 46    | 39.7        | 186    | 5     | US-10-425-115-55253  |

|    |    |      |     |   |                      |                   |
|----|----|------|-----|---|----------------------|-------------------|
| 28 | 46 | 39.7 | 232 | 5 | US-10-732-923-18914  | Sequence 18914, A |
| 29 | 46 | 39.7 | 278 | 4 | US-10-029-386-32852  | Sequence 32852, A |
| 30 | 46 | 39.7 | 319 | 4 | US-10-369-493-23683  | Sequence 23683, A |
| 31 | 46 | 39.7 | 319 | 4 | US-10-282-122A-43119 | Sequence 43119, A |
| 32 | 46 | 39.7 | 440 | 4 | US-10-094-749-1912   | Sequence 1912, Ap |
| 33 | 46 | 39.7 | 523 | 5 | US-10-450-763-55252  | Sequence 55252, A |
| 34 | 46 | 39.7 | 546 | 4 | US-10-112-944-712    | Sequence 712, App |
| 35 | 46 | 39.7 | 546 | 5 | US-10-450-763-40380  | Sequence 40380, A |
| 36 | 46 | 39.7 | 599 | 3 | US-09-847-102A-37    | Sequence 37, Appl |
| 37 | 46 | 39.7 | 682 | 4 | US-10-152-548-16     | Sequence 16, Appl |
| 38 | 46 | 39.7 | 685 | 3 | US-09-847-102A-57    | Sequence 57, Appl |
| 39 | 46 | 39.7 | 685 | 4 | US-10-146-474-14     | Sequence 14, Appl |
| 40 | 46 | 39.7 | 685 | 4 | US-10-301-764-14     | Sequence 14, Appl |
| 41 | 46 | 39.7 | 685 | 5 | US-10-882-586A-9     | Sequence 9, Appl  |
| 42 | 46 | 39.7 | 694 | 3 | US-09-847-102A-56    | Sequence 56, Appl |
| 43 | 46 | 39.7 | 694 | 4 | US-10-205-823-144    | Sequence 144, App |
| 44 | 46 | 39.7 | 694 | 4 | US-10-285-976-53     | Sequence 53, Appl |
| 45 | 46 | 39.7 | 694 | 4 | US-10-295-027-1344   | Sequence 1344, Ap |

## ALIGNMENTS

### RESULT 1

US-10-383-930-22  
; Sequence 22, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 22  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-22

Query Match 100.0%; Score 116; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 4.4e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEANYAGNQSIGNYRGWFNP 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 QEANYAGNQSIGNYRGWFNP 20

### RESULT 2

US-10-797-821-22  
; Sequence 22, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049

```
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 22
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: GspB peptide
US-10-797-821-22

Query Match          100.0%; Score 116; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.4e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEANYAGNQSIGNYRGWFNP 20
   |||||
Db 1 QEANYAGNQSIGNYRGWFNP 20
   |||||

RESULT 3
US-10-383-930-29
; Sequence 29, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: US/10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-29

Query Match          100.0%; Score 116; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEANYAGNQSIGNYRGWFNP 20
   |||||
Db 403 QEANYAGNQSIGNYRGWFNP 422
   |||||

RESULT 4
US-10-383-930-30
; Sequence 30, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-29

Query Match          100.0%; Score 116; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEANYAGNQSIGNYRGWFNP 20
   |||||
Db 403 QEANYAGNQSIGNYRGWFNP 422
   |||||

RESULT 5
US-10-383-930-33
; Sequence 33, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: US/10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-33

Query Match          100.0%; Score 116; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEANYAGNQSIGNYRGWFNP 20
   |||||
Db 403 QEANYAGNQSIGNYRGWFNP 422
   |||||

RESULT 6
US-10-797-821-29
; Sequence 29, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: US/10/797,821
; PRIOR FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
```

```
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-29

Query Match      100.0%; Score 116; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QEANYAGNQSIGNYRGWFNP 20
      |||||||
Db      403 QEANYAGNQSIGNYRGWFNP 422

RESULT 7
US-10-797-821-30
; Sequence 30, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-30

Query Match      100.0%; Score 116; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QEANYAGNQSIGNYRGWFNP 20
      |||||||
Db      403 QEANYAGNQSIGNYRGWFNP 422

RESULT 8
US-10-797-821-33
; Sequence 33, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
```

```
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-33

Query Match      100.0%; Score 116; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QEANYAGNQSIGNYRGWFNP 20
      |||||||
Db      403 QEANYAGNQSIGNYRGWFNP 422

RESULT 9
US-10-383-930-31
; Sequence 31, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-31

Query Match      100.0%; Score 116; DB 4; Length 432;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QEANYAGNQSIGNYRGWFNP 20
      |||||||
Db      404 QEANYAGNQSIGNYRGWFNP 423

RESULT 10
US-10-383-930-32
; Sequence 32, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; LENGTH: 432
; TYPE: PRT
```

```
; ORGANISM: Streptococcus mutans
US-10-383-930-32

Query Match      100.0%; Score 116; DB 4; Length 432;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEANYAGNQSIGNYRGWFNP 20
Db 404 QEANYAGNQSIGNYRGWFNP 423

RESULT 11
US-10-797-821-31
; Sequence 31, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-31

Query Match      100.0%; Score 116; DB 5; Length 432;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEANYAGNQSIGNYRGWFNP 20
Db 404 QEANYAGNQSIGNYRGWFNP 423

RESULT 12
US-10-797-821-32
; Sequence 32, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
```

```
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-32

Query Match      100.0%; Score 116; DB 5; Length 432;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEANYAGNQSIGNYRGWFNP 20
Db 404 QEANYAGNQSIGNYRGWFNP 423

RESULT 13
US-10-472-928-4652
; Sequence 4652, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4652
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: secreted 45 kd protein (usp45)
; OTHER INFORMATION: Cellular location: outside
; OTHER INFORMATION: Similar to strain R6 sequence 15904062 (0.E+01)
US-10-472-928-4652

Query Match      82.8%; Score 96; DB 5; Length 392;
Best Local Similarity 78.9%; Pred. No. 7.3e-06;
Matches 15; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EANYAGNQSIGNYRGWFNP 20
Db 360 ESNYAGNRTIGNHRGWFP 378

RESULT 14
US-10-617-320-3230
; Sequence 3230, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
```

Search completed: February 15, 2006, 22:39:15  
Job time : 166 secs

OPERATING SYSTEM: <Unknown>  
SOFTWARE: <Unknown>  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/617,320  
FILING DATE: 10-Jul-2003  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,433  
FILING DATE: 30-Jun-1998  
APPLICATION NUMBER: 60/ 085131  
FILING DATE: May 12, 1998  
APPLICATION NUMBER: 60/051553  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 3230:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 399 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...399  
SEQUENCE DESCRIPTION: SEQ ID NO: 3230:  
US-10-617-320-3230

Query Match 82.8%; Score 96; DB 5; Length 399;  
Best Local Similarity 78.9%; Pred. No. 7.4e-06;  
Matches 15; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Oy 2 EANYAGNOSIGNRGWFNP 20  
|:|||||:|||||:  
Db 367 ESNYAGNRTIGNRGWFNP 385

RESULT 15  
US-10-474-792-600  
; Sequence 600, Application US/10474792  
; Publication No. US20040236072A1  
; GENERAL INFORMATION:  
; APPLICANT: Olmsted, Stephen  
; APPLICANT: Zagursky, Robert  
; APPLICANT: Nickbarg, Elliot  
; APPLICANT: Winter, Lourie  
; TITLE OF INVENTION: SURFACE PROTEINS OF STREPTOCOCCUS PYOGENES  
; FILE REFERENCE: AM 100399  
; CURRENT APPLICATION NUMBER: US/10/474,792  
; CURRENT FILING DATE: 2003-10-14  
; NUMBER OF SEQ ID NOS: 674  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 600  
; LENGTH: 398  
; TYPE: PRT  
; ORGANISM: Streptococcus pyogenes  
US-10-474-792-600

Query Match 66.4%; Score 77; DB 5; Length 398;  
Best Local Similarity 63.2%; Pred. No. 0.0049;  
Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Oy 2 EANYAGNOSIGNRGWFNP 20  
|:|||||:|||||:  
Db 371 ESNYGRQYIADHRGWFP 389

**THIS PAGE BLANK (USPTO)**

| Result No. | Score |        | Query |   | DB                 | ID                | Description |
|------------|-------|--------|-------|---|--------------------|-------------------|-------------|
|            | Match | Length | %     |   |                    |                   |             |
| 1          | 116   | 100.0  | 431   | 7 | US-11-052-554A-210 | Sequence 210, App |             |
| 2          | 77    | 66.4   | 398   | 7 | US-11-052-554A-452 | Sequence 358, App |             |
| 3          | 74    | 63.8   | 544   | 7 | US-11-052-554A-358 | Sequence 358, App |             |
| 4          | 70    | 60.3   | 211   | 7 | US-11-052-554A-352 | Sequence 352, App |             |
| 5          | 46    | 39.7   | 702   | 6 | US-10-510-386-214  | Sequence 214, App |             |
| 6          | 42    | 36.2   | 106   | 7 | US-11-004-590-390  | Sequence 390, App |             |
| 7          | 42    | 36.2   | 987   | 7 | US-11-052-554A-153 | Sequence 153, App |             |
| 8          | 41    | 35.3   | 138   | 7 | US-11-089-872-3    | Sequence 3, Appli |             |
| 9          | 41    | 35.3   | 213   | 7 | US-11-158-505-34   | Sequence 34, Appl |             |
| 10         | 41    | 35.3   | 794   | 6 | US-10-485-517-355  | Sequence 355, App |             |
| 11         | 40    | 34.5   | 108   | 6 | US-10-771-257-39   | Sequence 39, Appl |             |
| 12         | 40    | 34.5   | 108   | 7 | US-11-127-677-39   | Sequence 39, Appl |             |
| 13         | 40    | 34.5   | 3157  | 7 | US-11-052-554A-142 | Sequence 142, App |             |
| 14         | 39.5  | 34.1   | 227   | 7 | US-11-170-653-27   | Sequence 27, Appl |             |
| 15         | 39.5  | 34.1   | 3300  | 7 | US-11-052-554A-133 | Sequence 133, App |             |
| 16         | 39    | 33.6   | 94    | 7 | US-11-097-812-9    | Sequence 9, Appli |             |
| 17         | 39    | 33.6   | 106   | 6 | US-10-485-517-181  | Sequence 181, App |             |
| 18         | 39    | 33.6   | 106   | 7 | US-11-004-590-388  | Sequence 388, App |             |
| 19         | 39    | 33.6   | 107   | 7 | US-11-064-174-173  | Sequence 173, App |             |
| 20         | 39    | 33.6   | 107   | 7 | US-11-064-174-179  | Sequence 179, App |             |
| 21         | 39    | 33.6   | 107   | 7 | US-11-154-337-5    | Sequence 5, Appli |             |
| 22         | 39    | 33.6   | 107   | 7 | US-11-182-908-5    | Sequence 5, Appli |             |
| 23         | 39    | 33.6   | 108   | 6 | US-10-665-658-3    | Sequence 3, Appli |             |
| 24         | 39    | 33.6   | 108   | 6 | US-10-834-397-14   | Sequence 14, Appl |             |
| 25         | 39    | 33.6   | 108   | 7 | US-11-120-338-3    | Sequence 3, Appli |             |

```
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 252
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes MGAS8232
US-11-052-554A-252

Query Match      66.4%; Score 77; DB 7; Length 398;
Best Local Similarity 63.2%; Pred. No. 0.00018;
Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFNP 20
Db 371 ESNYSGROYIADHRGWFP 389

RESULT 3
US-11-052-554A-358
; Sequence 358, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 358
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-358

Query Match      63.8%; Score 74; DB 7; Length 544;
Best Local Similarity 60.0%; Pred. No. 0.00068;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 QEANYAGNOSIGNYRGWFNP 20
Db 511 KEANYKNOQYISNFRGWFP 530

RESULT 4
US-11-052-554A-352
; Sequence 352, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 352
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-352

Query Match      60.3%; Score 70; DB 7; Length 211;
Best Local Similarity 57.9%; Pred. No. 0.00011;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
```

```
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 EANYAGNOSIGNYRGWFNP 20
Db 178 ESNYKDOQWVDNYRGWFDP 196

RESULT 5
US-10-510-386-214
; Sequence 214, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 214
; LENGTH: 702
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-10-510-386-214

Query Match      39.7%; Score 46; DB 6; Length 702;
Best Local Similarity 43.5%; Pred. No. 15;
Matches 10; Conservative 3; Mismatches 6; Indels 4; Gaps 1;

QY 1 QEANYAGNOSIGNYR----GWFN 19
Db 435 QELSPAGGVKKGYRGNTIGWAN 457

RESULT 6
US-11-004-590-390
; Sequence 390, Application US/11004590
; Publication No. US20060008883A1
; GENERAL INFORMATION:
; APPLICANT: Lazar, Gregory Alan
; APPLICANT: Desjarlais, John R.
; TITLE OF INVENTION: METHODS OF GENERATING VARIANT PROTEINS WITH INCREASED HOST STRING
; FILE OF INVENTION: CONTENT AND COMPOSITIONS THEREOF
; FILE REFERENCE: 185832/US/5
; CURRENT APPLICATION NUMBER: US/11/004,590
; CURRENT FILING DATE: 2004-12-03
; PRIOR APPLICATION NUMBER: US 60/527,167
; PRIOR FILING DATE: 2003-12-04
; PRIOR APPLICATION NUMBER: US 60/581,613
; PRIOR FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US 60/601,665
; PRIOR FILING DATE: 2004-08-13
; PRIOR APPLICATION NUMBER: US 60/619,483
; NUMBER OF SEQ ID NOS: 458
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 390
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-004-590-390

Query Match      36.2%; Score 42; DB 7; Length 106;
Best Local Similarity 54.5%; Pred. No. 9.1;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 8 NQSIGNYRGWF 18
:|:|:|:|:|:|:
```



Db 26 SQNNYLGWY 36

RESULT 7

US-11-052-554A-153

; Sequence 153, Application US/11052554A

; Publication No. US20050288866A1

; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

; FILE REFERENCE: 30853/40359A

; CURRENT APPLICATION NUMBER: US/11/052,554A

; CURRENT FILING DATE: 2005-02-07

; PRIOR APPLICATION NUMBER: US 60/589,227

; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; PRIOR FILING DATE: 2004-02-06

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 153

; LENGTH: 987

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis H37RV

US-11-052-554A-153

Query Match 36.2%; Score 42; DB 7; Length 987;

Best Local Similarity 42.9%; Pred. No. 82;

Matches 9; Conservative 0; Mismatches 8; Indels 4; Gaps 1;

QY 3 ANYAGNQSIGN----YRGWFN 19

Db 901 AGIGGSGFGSGSLSGWFN 921

RESULT 8

US-11-089-872-3

; Sequence 3, Application US/11089872

; Publication No. US20050260212A1

; GENERAL INFORMATION:

; APPLICANT: Zhang, Yongke

; APPLICANT: Ramakrishnan, Vanitha

; APPLICANT: Law, Debbie

; TITLE OF INVENTION: ANTI-LFL2 ANTIBODIES FOR THE DIAGNOSIS, PROGNOSIS AND TREATMENT

; TITLE OF INVENTION: OF CANCER

; FILE REFERENCE: 05882.0223.NPUS02

; CURRENT APPLICATION NUMBER: US/11/089,872

; CURRENT FILING DATE: 2005-03-25

; PRIOR APPLICATION NUMBER: 60/557,440

; PRIOR FILING DATE: 2004-03-26

; PRIOR APPLICATION NUMBER: 60/638,708

; PRIOR FILING DATE: 2004-12-22

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 3

; LENGTH: 138

; TYPE: PRT

; ORGANISM: Mus musculus

US-11-089-872-3

Query Match 35.3%; Score 41; DB 7; Length 138;

Best Local Similarity 85.7%; Pred. No. 17;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 GNYRGWGF 18

Db 120 GNYRAWF 126

RESULT 9

US-11-158-505-34

; Sequence 34, Application US/11158505

; Publication No. US20060002921A1

; GENERAL INFORMATION:

; APPLICANT: WINSOR-HINES, DAWN

; APPLICANT: RAO, PATRICIA

; APPLICANT: RINGLER, DOUGLAS J

; APPLICANT: PONATH, PAUL

; TITLE OF INVENTION: OPTIMIZED DOSING OF ANTI-CD4 ANTIBODIES FOR TOLERANCE

; TITLE OF INVENTION: INDUCTION IN PRIMATES

; FILE REFERENCE: TLN-031

; CURRENT APPLICATION NUMBER: US/11/158,505

; CURRENT FILING DATE: 2005-06-21

; PRIOR APPLICATION NUMBER: 60/582,181

; PRIOR FILING DATE: 2004-06-22

; NUMBER OF SEQ ID NOS: 76

; SOFTWARE: PatentIn Ver. 3.3

; SEQ ID NO 34

; LENGTH: 213

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Humanized CD8

; OTHER INFORMATION: antibody light chain construct

US-11-158-505-34

Query Match 35.3%; Score 41; DB 7; Length 213;

Best Local Similarity 50.0%; Pred. No. 26;

Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 7 GNOQSIGNYRGWF 18

Db 25 GSQDINNYLAWY 36

RESULT 10

US-10-485-517-355

; Sequence 355, Application US/10485517

; Publication No. US20050256299A1

; GENERAL INFORMATION:

; APPLICANT: University of Sheffield

; APPLICANT: Biosynex Incorporated

; APPLICANT: Foster, Simon

; APPLICANT: Mond, James

; TITLE OF INVENTION: Antigenic Polypeptides

; FILE REFERENCE: PI00629WO

; CURRENT APPLICATION NUMBER: US/10/485,517

; CURRENT FILING DATE: 2004-02-02

; PRIOR APPLICATION NUMBER: GB 0118825.9

; PRIOR FILING DATE: 2001-08-02

; PRIOR APPLICATION NUMBER: GB 0200349.9

; PRIOR FILING DATE: 2002-01-09

; NUMBER OF SEQ ID NOS: 424

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 355

; LENGTH: 794

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-10-485-517-355

Query Match 35.3%; Score 41; DB 6; Length 794;

Best Local Similarity 100.0%; Pred. No. 93;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 YRGWFN 19

Db 562 YRGWFN 567

RESULT 11

US-10-771-257-39

; Sequence 39, Application US/10771257

; Publication No. US20050288864A1

; GENERAL INFORMATION:

; APPLICANT: Medical Research Council

; APPLICANT: SISSA - Scuola Superiore Internazionale di Studi Avanzati

APPLICANT: Cattaneo, Antonino  
APPLICANT: Maritan, Amos  
APPLICANT: Visintin, Michela  
APPLICANT: Rabbitts, Terrence H  
APPLICANT: Settanni, Giovanni  
TITLE OF INVENTION: Intracellular antibodies  
FILE REFERENCE: 18396/2272  
CURRENT APPLICATION NUMBER: US/10/771,257  
CURRENT FILING DATE: 2004-02-03  
PRIOR APPLICATION NUMBER: PCT/GB02/03512  
PRIOR FILING DATE: 2002-08-01  
PRIOR APPLICATION NUMBER: GB 0119004.0  
PRIOR FILING DATE: 2001-08-03  
PRIOR APPLICATION NUMBER: GB 0121577.1  
PRIOR FILING DATE: 2001-09-06  
PRIOR APPLICATION NUMBER: GB 0200928.0  
PRIOR FILING DATE: 2002-01-16  
PRIOR APPLICATION NUMBER: GB 0203569.9  
PRIOR FILING DATE: 2002-02-14  
PRIOR APPLICATION NUMBER: IT RM2001A000633  
PRIOR FILING DATE: 2001-10-25  
NUMBER OF SEQ ID NOS: 124  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 39  
LENGTH: 108  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-771-257-39

Query Match 34.5%; Score 40; DB 6; Length 108;  
Best Local Similarity 54.5%; Pred. No. 18;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 8 NQSIGNRGWF 18  
:||||:|  
Db 26 SQSIGSYLNWY 36

RESULT 12  
US-11-127-677-39  
Sequence 39, Application US/11127677  
Publication No. US20050272107A1  
GENERAL INFORMATION:  
APPLICANT: Medical Research Council  
APPLICANT: Rabbitts, Terrence H  
APPLICANT: Tanaka, Tomoyuki  
TITLE OF INVENTION: Intracellular antibodies  
FILE REFERENCE: 18396/2462  
CURRENT APPLICATION NUMBER: US/11/127,677  
CURRENT FILING DATE: 2005-05-12  
PRIOR APPLICATION NUMBER: PCT/GB03/04942  
PRIOR FILING DATE: 2003-11-14  
PRIOR APPLICATION NUMBER: GB 0226729.2  
PRIOR FILING DATE: 2002-11-15  
NUMBER OF SEQ ID NOS: 150  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 39  
LENGTH: 108  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Derived protein sequence of scFv  
US-11-127-677-39

Query Match 34.5%; Score 40; DB 7; Length 108;  
Best Local Similarity 54.5%; Pred. No. 18;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 8 NQSIGNRGWF 18  
:||||:|  
Db 26 SQSIGSYLNWY 36

RESULT 13  
US-11-052-554A-142  
Sequence 142, Application US/11052554A  
Publication No. US20050288866A1  
GENERAL INFORMATION:  
APPLICANT: Sachdeva, et al.  
TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
FILE REFERENCE: 30853/40359A  
CURRENT APPLICATION NUMBER: US/11/052.554A  
CURRENT FILING DATE: 2005-02-07  
PRIOR APPLICATION NUMBER: US 60/589,227  
PRIOR FILING DATE: 2004-07-20  
PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
PRIOR FILING DATE: 2004-02-06  
NUMBER OF SEQ ID NOS: 763  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 142  
LENGTH: 3157  
TYPE: PRT  
ORGANISM: Mycobacterium tuberculosis H37Rv  
US-11-052-554A-142

Query Match 34.5%; Score 40; DB 7; Length 3157;  
Best Local Similarity 40.0%; Pred. No. 5.1e+02;  
Matches 8; Conservative 2; Mismatches 4; Indels 6; Gaps 1;

QY 6 AGNOSIGNYR-----GWFN 19  
|||:|:  
Db 1575 AGNTNTGGFNGFNVNTGWFN 1594

RESULT 14  
US-11-170-653-27  
Sequence 27, Application US/11170653  
Publication No. US20050271769A1  
GENERAL INFORMATION:  
APPLICANT: Danisco A/S  
APPLICANT: Sibbesen, Ole  
APPLICANT: Sorensen, Jens  
TITLE OF INVENTION: Xylanase Variants Having Altered Sensitivity to Xylanase Inhibitors  
FILE REFERENCE: 674509-2046  
CURRENT APPLICATION NUMBER: US/11/170,653  
CURRENT FILING DATE: 2005-06-23  
PRIOR APPLICATION NUMBER: US/10/237,386  
PRIOR FILING DATE: 2002-09-09  
PRIOR APPLICATION NUMBER: PCT/IB01/00426  
PRIOR FILING DATE: 2001-03-08  
PRIOR APPLICATION NUMBER: GB 0005585.5  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: GB 0015751.1  
PRIOR FILING DATE: 2000-06-27  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 27  
LENGTH: 227  
TYPE: PRT  
ORGANISM: H. insolens  
US-11-170-653-27

Query Match 34.1%; Score 39.5; DB 7; Length 227;  
Best Local Similarity 41.7%; Pred. No. 46;  
Matches 10; Conservative 2; Mismatches 5; Indels 7; Gaps 2;

QY 4 NYAG-----NQSIG---NYRGWFNP 20  
|:|:|:|:|  
Db 80 NFVGKGNPCTGRTINYGYFNP 103

RESULT 15  
US-11-052-554A-133  
Sequence 133, Application US/11052554A  
Publication No. US20050288866A1

```
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 133
; LENGTH: 3300
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-133
```

```
Query Match      34.1%; Score 39.5; DB 7; Length 3300;
Best Local Similarity 43.5%; Pred. No. 6.4e+02;
Matches 10; Conservative 1; Mismatches 5; Indels 7; Gaps 1;
```

```
QY      4 NYAGNQSIGNVR-----GWFN 19
Db      544 NILGNANIGNYILGSGNVGDFN 566
```

Search completed: February 15, 2006, 22:39:38  
Job time : 18 secs

2006 FEB 16 08:44:51

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 17:57:02 ; Search time 113.754 Seconds  
(without alignments)  
1668.609 Million cell updates/sec

Title: US-10-797-821-31

Perfect score: 2110

Sequence: 1 MKKRILSAVLNVTGLSSAT.....SIGNVRGWNFGSVSYIPN 432

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_21.\*  
1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003bs.\*  
8: Geneseq2004s.\*  
9: Geneseq2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID       | Description         |
|------------|--------|-------------|--------|-------------|---------------------|
| 1          | 2110   | 100.0       | 432    | 7 ADD93651  | Add93651 Streptoco  |
| 2          | 2110   | 100.0       | 432    | 9 ADX37274  | Adx37274 Streptoco  |
| 3          | 2099   | 99.5        | 432    | 7 ADD93652  | Add93652 Streptoco  |
| 4          | 2099   | 99.5        | 432    | 9 ADX37275  | Adx37275 Streptoco  |
| 5          | 2083.5 | 98.7        | 431    | 7 ADD93650  | Add93650 Streptoco  |
| 6          | 2083.5 | 98.7        | 431    | 9 ADX37273  | Adx37273 Streptoco  |
| 7          | 2079.5 | 98.6        | 431    | 9 ADD93649  | Add93649 Streptoco  |
| 8          | 2079.5 | 98.6        | 431    | 9 ADX37272  | Adx37272 Streptoco  |
| 9          | 2079.5 | 98.6        | 431    | 9 AEB91500  | Aeb91500 Microbial  |
| 10         | 2075.5 | 98.4        | 431    | 7 ADD93653  | Add93653 Streptoco  |
| 11         | 2075.5 | 98.4        | 431    | 9 ADX37276  | Adx37276 Streptoco  |
| 12         | 1234.5 | 58.5        | 447    | 8 ABP29684  | Abp29684 Streptoco  |
| 13         | 1234.5 | 58.5        | 447    | 8 ADU69524  | Adu69524 S agalact  |
| 14         | 1234.5 | 58.5        | 447    | 8 ADV881808 | Adv881808 Streptoco |
| 15         | 1234.5 | 58.5        | 447    | 8 ADV79645  | Adv79645 Streptoco  |
| 16         | 1234.5 | 58.5        | 447    | 8 AEB91542  | Aeb91542 Microbial  |
| 17         | 1093   | 51.8        | 398    | 9 AEB25919  | Abp25919 Streptoco  |
| 18         | 1087   | 51.5        | 398    | 8 ADR38884  | Adr38884 S. pyogen  |
| 19         | 1087   | 51.5        | 398    | 8 AEB25918  | Abp25918 Streptoco  |
| 20         | 1077.5 | 51.1        | 395    | 5 ABP25917  | Abp25917 Streptoco  |
| 21         | 1059   | 50.2        | 392    | 6 ABU02747  | Abu02747 S. pneumo  |
| 22         | 1059   | 50.2        | 392    | 8 ADT50226  | Adt50226 S. pneumo  |
| 23         | 1059   | 50.2        | 392    | 8 ADT50165  | Adt50165 S. pneumo  |
| 24         | 1059   | 50.2        | 399    | 8 ADR94595  | Adr94595 Novel S.   |

## ALIGNMENTS

RESULT 1

ADD93651

ID ADD93651 standard; protein; 432 AA.

XX AC ADD93651;

XX XX

DT 29-JAN-2004 (first entry)

XX XX

DE Streptococcus mutans glucan binding protein-B.

XX XX

KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.

XX OS

OS Streptococcus mutans.

XX XX

PN WO2003075845-A2.

XX PD

PD 18-SEP-2003.

XX XX

PF 07-MAR-2003; 2003WO-US006962.

XX XX

PR 07-MAR-2002; 2002US-0363209P.

XX PR

PR 08-AUG-2002; 2002US-0402483P.

XX XX

(FORS-) FORSYTH INST.

PA PA

XX Smith DJ, Taubman MA;

PI PI

XX WPI; 2003-845091/78.

DR DR

XX GENBANK; AY046412.

XX XX

PT Composition useful as vaccines for dental caries comprises a fragment of

XX PT

XX a glucan binding protein-B binding to a major histocompatibility complex

XX PT

XX class II protein.

XX XX

PS Claim 5; Page 8; 49pp; English.

XX XX

CC The present sequence is the protein sequence of the glucan binding

CC protein-B (GbpB) of Streptococcus mutans strain 15p2. The sequence

CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The

CC invention provides immunogenic compositions and vaccines for dental

CC caries. The compositions comprise major histocompatibility complex (MHC)

CC class II protein-binding GbpB peptides, especially HLA-binding peptides,

CC covalently linked with peptide subunits of a glucosyltransferase. The

CC compositions are used in a claimed method of eliciting production of an

CC antibody in a mammal. Dieptopic or multiptopic polypeptides can be

CC prepared synthetically or by recombinant DNA technology. Antibodies

CC raised against MHC class II binding fragments of GbpB can be used in

Aea58465 Streptoco  
Adk47859 Streptoco  
Adt50227 S.pneumon  
Aar14150 MSP encod  
Aar14530 Usp45 pro  
Abb55584 Lactococc  
Abu29734 Protein e  
Adc95468 E. faeciu  
Aay22579 Bacterial  
Adv16553 E. faecal  
Adh88105 Enterococ  
Aay00250 Enterococ  
Abp43469 E faecali  
Abu88497 E. faecal  
Abu13748 Enterococ  
Adv16734 E. faecal  
Ady39216 Novel Ent  
Adh87829 Enterococ  
Aay00251 Enterococ  
Abp43470 E faecali  
Abu88498 E. faecal

399 9 AEA58465  
392 8 ADK47859  
392 8 ADT50227  
461 2 AAR14150  
461 2 AAR14530  
456 5 ABB55584  
524 6 ABU29734  
525 7 ADC95468  
210 2 AAY22579  
482 9 ADV16553  
497 7 ADH88105  
449 2 AAY00250  
449 5 ABP43469  
449 6 ABU88497  
449 6 ABU13748  
449 9 ADV16734  
449 9 ADY39216  
450 7 ADH87829  
422 2 AAY00251  
444 5 ABP43470  
422 6 ABU88498

25 1059 50.2  
26 1058 50.1  
27 1058 50.1  
28 681.5 32.3  
29 680.5 32.3  
30 660 31.3  
31 588.5 27.9  
32 584.5 27.7  
33 561 26.6  
34 529 25.1  
35 529 25.1  
36 500.5 23.7  
37 500.5 23.7  
38 500.5 23.7  
39 500.5 23.7  
40 500.5 23.7  
41 500.5 23.7  
42 500.5 23.7  
43 452.5 21.4  
44 452.5 21.4  
45 452.5 21.4

CC passive immunisation.  
 XX Sequence 432 AA;  
 SQ

Query Match 100.0%; Score 2110; DB 7; Length 432;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-134;  
 Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAIAKDDFDAQIASQDSKINNLTAAQQAQAQVNTIIG 60  
 DB 1 MKKRILSAVLVSGVTLSSTATTLSAIAKDDFDAQIASQDSKINNLTAAQQAQAQVNTIIG 60  
 QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAOKSNAATSY 120  
 DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAOKSNAATSY 120  
 QY 121 INAIINSKVSDAINRVSAIREVVSANERKMLQQEQDKAAVEQKQENQAAINTVAANO 180  
 DB 121 INAIINSKVSDAINRVSAIREVVSANERKMLQQEQDKAAVEQKQENQAAINTVAANO 180  
 QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQQDKATLVAQKAAAEAEARQAAAAA 240  
 DB 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQQDKATLVAQKAAAEAEARQAAAAA 240  
 QY 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300  
 DB 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300  
 QY 301 SDOESTTATAAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWGGGOWA 360  
 DB 301 SDOESTTATAAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWGGGOWA 360  
 QY 361 ASAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTGVGGQIQVQEAANYAGNQSIGNYRGW 420  
 DB 361 ASAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTGVGGQIQVQEAANYAGNQSIGNYRGW 420  
 QY 421 FNPGSVSYIYPN 432  
 DB 421 FNPGSVSYIYPN 432

RESULT 2  
 ADX37274  
 ID ADX37274 standard; protein; 432 AA.  
 XX ADX37274;  
 AC ADX37274;  
 XX  
 DT 21-APR-2005 (first entry)  
 XX Streptococcus mutant glucan binding protein B variant #3.  
 DE immunogenicity; immune stimulation; glucan binding protein-B;  
 KW microparticle; major histocompatibility complex; tooth disease.  
 XX Streptococcus mutans.  
 OS  
 PN US2005031633-A1.  
 XX  
 PD 10-FEB-2005.  
 XX  
 PF 09-MAR-2004; 2004US-00797821.  
 XX  
 PR 13-APR-1998; 98US-0081550P.  
 PR 08-JAN-1999; 99US-0115142P.  
 PR 12-APR-1999; 99US-0029004P.  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 PR 07-MAR-2003; 2003US-00383930.  
 XX (SMIT/) SMITH D J.  
 PA (TAUB/) TAUBMAN M A.  
 XX Smith DJ, Taubman MA;  
 PI

XX WPI; 2005-151644/16.  
 DR  
 XX New composition comprising a fragment of a glucan binding protein-B  
 PT (GbpB) that binds to MHC class II protein, and a biocompatible  
 PT microparticle, useful for producing an antibody (claimed) for immunizing  
 PT mammals against dental caries.  
 XX  
 PS Claim 3; SEQ ID NO 31; 73pp; English.  
 XX  
 CC The invention relates to a composition comprising a fragment of a glucan  
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
 CC fragment binds to a major histocompatibility complex (MHC) class II  
 CC protein. The composition is useful for producing an antibody for  
 CC immunizing mammals against dental caries. This sequence corresponds to a  
 CC Streptococcus mutans GbpB protein of the invention.  
 XX  
 SQ Sequence 432 AA;  
 Query Match 100.0%; Score 2110; DB 9; Length 432;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-134;  
 Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAIAKDDFDAQIASQDSKINNLTAAQQAQAQVNTIIG 60  
 DB 1 MKKRILSAVLVSGVTLSSTATTLSAIAKDDFDAQIASQDSKINNLTAAQQAQAQVNTIIG 60  
 QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAOKSNAATSY 120  
 DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAOKSNAATSY 120  
 QY 121 INAIINSKVSDAINRVSAIREVVSANERKMLQQEQDKAAVEQKQENQAAINTVAANO 180  
 DB 121 INAIINSKVSDAINRVSAIREVVSANERKMLQQEQDKAAVEQKQENQAAINTVAANO 180  
 QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQQDKATLVAQKAAAEAEARQAAAAA 240  
 DB 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQQDKATLVAQKAAAEAEARQAAAAA 240  
 QY 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300  
 DB 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300  
 QY 301 SDOESTTATAAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWGGGOWA 360  
 DB 301 SDOESTTATAAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWGGGOWA 360  
 QY 361 ASAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTGVGGQIQVQEAANYAGNQSIGNYRGW 420  
 DB 361 ASAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTGVGGQIQVQEAANYAGNQSIGNYRGW 420  
 QY 421 FNPGSVSYIYPN 432  
 DB 421 FNPGSVSYIYPN 432

RESULT 3  
 ADD93652  
 ID ADD93652 standard; protein; 432 AA.  
 XX ADD93652;  
 AC ADD93652;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX Streptococcus mutans glucan binding protein-B.  
 DE Streptococcus mutans glucan binding protein-B.  
 KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.  
 XX Streptococcus mutans.  
 OS  
 PN WO2003075845-A2.  
 XX  
 PD 18-SEP-2003.

XX 07-MAR-2003; 2003WO-US006962.  
XX PF  
XX 07-MAR-2002; 2002US-0363209P.  
XX PR  
XX 08-AUG-2002; 2002US-0402483P.  
XX PA  
XX (FORS-) FORSYTH INST.  
XX Smith DJ, Taubman MA;  
XX WPI: 2003-845091/78.  
XX GENBANK; AY046413.  
XX  
XX Composition useful as vaccines for dental caries comprises a fragment of  
XX PT a glucan binding protein-B binding to a major histocompatibility complex  
XX PT class II protein.  
XX  
XX Claim 5; Page 8; 49pp; English.  
XX  
XX The present sequence is the protein sequence of the glucan binding  
XX CC protein-B (GbpB) of Streptococcus mutans strain 3SM1. The sequence  
XX CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
XX CC invention provides immunogenic compositions and vaccines for dental  
XX CC caries. The compositions comprise major histocompatibility complex (MHC)  
XX CC class II protein-binding GbpB peptides, especially HLA-binding peptides,  
XX CC covalently linked with peptide subunits of a glucosyltransferase. The  
XX CC compositions are used in a claimed method of eliciting production of an  
XX CC antibody in a mammal. Diepitopic or multiepitopic polypeptides can be  
XX CC prepared synthetically or by recombinant DNA technology. Antibodies  
XX CC raised against MHC class II binding fragments of GbpB can be used in  
XX CC passive immunisation.  
XX  
XX Sequence 432 AA;

Query Match 99.5%; Score 2099; DB 7; Length 432;  
Best Local Similarity 99.3%; Pred. No. 3.2e-133;  
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSTLSSATLSAIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60  
DB 1 MKKRILSAVLVSGVTLSSTLSSATLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60  
QY 61 QVSALQTOQAEIQAENRLEAQSATLGGQIQTLSKIVARNESLKKQARSQAQSNAAATSY 120  
DB 61 QVSALQTOQAEIQAENRLEAQSATLGGQIQTLSKIVARNESLKKQARSQAQSNAAATSY 120  
QY 121 INAIINSKSVSDAINRVSAIREVVSANEXMLQOEODKAAVEKQKQENQAANTVAANOE 180  
DB 121 INAIINSKSVSDAINRVSAIREVVSANEXMLHQEQDKAAVEKQKQENQAANTVAANOE 180  
QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEITTAQDOKATLVAKAAAEAAARQAAQAQAAEA 240  
DB 181 TTAQNTNALNTQQAQLEAAQLNLQAEITTAQDOKATLVAKAAAEAAARQAAQAQAAEA 240  
QY 241 KAAAEAKALQEAQAQAQAANNNNTQATDASDQQAADNTQAAQTGSDTDSAAQAVNN 300  
DB 241 KAAAEAKALQEAQAQAQAANNNNTQATDASDQQAADNTQAAQTGSDTDSAAQAVNN 300  
QY 301 SQESTTATAQPSASSASTAAVAANTSSANTYPAGCCTGWKSLAPWVGNWGGGWA 360  
DB 301 SQESTTATAQPSASSASTAAVAANTSSANTYPAGCCTGWKSLAPWVGNWGGGWA 360  
QY 361 ASAAAAGYRVGSTPSAGAVVNDGGYGHVAVYVTGQGGQIQVQENYAGNOSIGNYRGW 420  
DB 361 ASAAAAGYRVGSTPSAGAVVNDGGYGHVAVYVTGQGGQIQVQENYAGNOSIGNYRGW 420  
QY 421 FNPGSVSYIYPN 432  
DB 421 FNPGSVSYIYPN 432

RESULT 4  
ADX37275

ID ADX37275 standard; protein; 432 AA.  
XX AC  
XX ADX37275;  
XX DT  
XX 21-APR-2005 (first entry)  
XX DE  
XX Streptococcus mutant glucan binding protein B variant #4.  
XX KW immunogenicity; immune stimulation; glucan binding protein-B;  
XX KW microparticle; major histocompatibility complex; tooth disease.  
XX OS Streptococcus mutans.  
XX PN US2005031633-A1.  
XX PD 10-FEB-2005.  
XX PF 09-MAR-2004; 2004US-00797821.  
XX PR 13-APR-1998; 98US-0081550P.  
XX PR 08-JAN-1999; 99US-0115142P.  
XX PR 12-APR-1999; 99US-00290049.  
XX PR 07-MAR-2002; 2002US-0363209P.  
XX PR 08-AUG-2002; 2002US-0402483P.  
XX PR 07-MAR-2003; 2003US-00383930.  
XX (SMIT/) SMITH D J.  
XX (TAUB/) TAUBMAN M A.  
XX Smith DJ, Taubman MA;  
XX WPI: 2005-151644/16.  
XX  
XX New composition comprising a fragment of a glucan binding protein-B  
XX PT (GbpB) that binds to MHC class II protein, and a biocompatible  
XX PT microparticle, useful for producing an antibody (claimed) for immunizing  
XX PT mammals against dental caries.  
XX  
XX Claim 3; SEQ ID NO 32; 73pp; English.  
XX  
XX The invention relates to a composition comprising a fragment of a glucan  
XX CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
XX CC fragment binds to a major histocompatibility complex (MHC) class II  
XX CC protein. The composition is useful for producing an antibody for  
XX CC immunizing mammals against dental caries. This sequence corresponds to a  
XX CC Streptococcus mutans GbpB protein of the invention.  
XX  
XX Sequence 432 AA;

Query Match 99.5%; Score 2099; DB 9; Length 432;  
Best Local Similarity 99.3%; Pred. No. 3.2e-133;  
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSTLSSATLSAIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60  
DB 1 MKKRILSAVLVSGVTLSSTLSSATLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60  
QY 61 QVSALQTOQAEIQAENRLEAQSATLGGQIQTLSKIVARNESLKKQARSQAQSNAAATSY 120  
DB 61 QVSALQTOQAEIQAENRLEAQSATLGGQIQTLSKIVARNESLKKQARSQAQSNAAATSY 120  
QY 121 INAIINSKSVSDAINRVSAIREVVSANEXMLQOEODKAAVEKQKQENQAANTVAANOE 180  
DB 121 INAIINSKSVSDAINRVSAIREVVSANEXMLHQEQDKAAVEKQKQENQAANTVAANOE 180  
QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEITTAQDOKATLVAKAAAEAAARQAAQAQAAEA 240  
DB 181 TTAQNTNALNTQQAQLEAAQLNLQAEITTAQDOKATLVAKAAAEAAARQAAQAQAAEA 240  
QY 241 KAAAEAKALQEAQAQAQAANNNNTQATDASDQQAADNTQAAQTGSDTDSAAQAVNN 300  
DB 241 KAAAEAKALQEAQAQAQAANNNNTQATDASDQQAADNTQAAQTGSDTDSAAQAVNN 300

QY 301 SDOESTTATAOPSSASSASTAAVAANTSSANTYPAGCTGKSLAPWVNTWGGGOWA 360  
 Db |||||  
 QY 301 SDOESTTATAOPSSASSASTAAVAANTSSANTYPAGCTGKSLAPWVNTWGGGOWA 360  
 Db |||||  
 QY 361 ASAAAAGYRVGTPSAGAVAVWVNDGGYGHVAVVTGVQGGQIQVQEAANYAGNOSIGNYRCW 420  
 Db |||||  
 QY 361 ASAAAAGYRVGTPSAGAVAVWVNDGGYGHVAVVTGVQGGQIQVQEAANYAGNOSIGNYRCW 420  
 Db |||||  
 QY 421 FNPGSVSYIYPN 432  
 Db |||||  
 QY 421 FNPGSVSYIYPN 432  
 Db |||||  
 RESULT 5  
 ADD93650  
 ID ADD93650 standard; protein; 431 AA.  
 AC ADD93650;  
 XX  
 DT 29-JAN-2004 (first entry)  
 DE Streptococcus mutans glucan binding protein-B.  
 XX Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.  
 KW Streptococcus mutans.  
 OS Streptococcus mutans.  
 PN WO2003075845-A2.  
 XX  
 PD 18-SEP-2003.  
 XX  
 PF 07-MAR-2003; 2003WO-US006962.  
 XX  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 XX  
 PA (FORS-) FORSYTH INST.  
 XX  
 PI Smith DJ, Taubman MA;  
 XX WPI; 2003-845091/78.  
 DR GENBANK; AY046411.  
 XX  
 PT Composition useful as vaccines for dental caries comprises a fragment of  
 PT a glucan binding protein-B binding to a major histocompatibility complex  
 PT class II protein.  
 XX  
 PS Claim 5; Page 8; 49pp; English.  
 XX  
 CC The present sequence is the protein sequence of the glucan binding  
 CC protein-B (GbpB) of Streptococcus mutans strain 3VF4. The sequence  
 CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
 CC invention provides immunogenic compositions and vaccines for dental  
 CC caries. The compositions comprise major histocompatibility complex (MHC)  
 CC class II protein-binding GbpB peptides, especially HLA-binding peptides,  
 CC covalently linked with peptide subunits of a glucosyltransferase. The  
 CC compositions are used in a claimed method of eliciting production of an  
 CC antibody in a mammal. Diepitopic or multiepitopic polypeptides can be  
 CC prepared synthetically or by recombinant DNA technology. Antibodies  
 CC raised against MHC class II binding fragments of GbpB can be used in  
 CC passive immunisation.  
 XX  
 SQ Sequence 431 AA;  
 Query Match 98.7%; Score 2083.5; DB 7; Length 431;  
 Best Local Similarity 99.1%; Pred. No. 3.6e-132;  
 Matches 428; Conservative 1; Mismatches 1; Indels 1; Gaps 1;  
 QY 1 MKKRILSAVLVSGVTLSSTATTLSA KADDFDAQIASQDSKINLTAQQAQAQVNTIQ 60  
 Db |||||  
 QY 1 MKKRILSAVLVSGVTLSSTATTLSA KADDFDAQIASQDSKINLTAQQAQAQVNTIQ 60  
 Db |||||  
 QY 61 QVSALQTOQAEIQAENRLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 120

Db |||||  
 QY 61 QVSALQTOQAEIQAENRLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 120  
 Db |||||  
 QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQODKAAVEKQKQENQAAINTVAANOE 180  
 Db |||||  
 QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQODKAAVEKQKQENQAAINTVAANOE 180  
 Db |||||  
 QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEITTAQDOKATLVAQKAAAEFAAQAQAAAE 240  
 Db |||||  
 QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEITTAQDOKATLVAQKAAAEFAAQAQAAAE 240  
 Db |||||  
 QY 241 KAAAEKALQEQAQAQAQAANNNTQATDASDQQAADNTQAAQTGSDTQSAQAQVNN 300  
 Db |||||  
 QY 241 KAAAEKALQEQAQAQAQAANNNTQATDASDQQAADNTQAAQTGSDTQSAQAQVNN 300  
 Db |||||  
 QY 301 SDOESTTATAOPSSASSASTAAVAANTSSANTYPAGCTGKSLAPWVNTWGGGOWA 360  
 Db |||||  
 QY 300 SDOESTTATAOPSSASSASTAAVAANTSSANTYPAGCTGKSLAPWVNTWGGGOWA 359  
 Db |||||  
 QY 361 ASAAAAGYRVGTPSAGAVAVWVNDGGYGHVAVVTGVQGGQIQVQEAANYAGNOSIGNYRCW 420  
 Db |||||  
 QY 360 ASAAAAGYRVGTPSAGAVAVWVNDGGYGHVAVVTGVQGGQIQVQEAANYAGNOSIGNYRCW 419  
 Db |||||  
 QY 421 FNPGSVSYIYPN 432  
 Db |||||  
 QY 420 FNPGSVSYIYPN 431  
 Db |||||  
 RESULT 6  
 ADX37273  
 ID ADX37273 standard; protein; 431 AA.  
 AC ADX37273;  
 XX  
 DT 21-APR-2005 (first entry)  
 DE Streptococcus mutant glucan binding protein B variant #2.  
 KW immunogenicity; immune stimulation; glucan binding protein-B;  
 KW microparticle; major histocompatibility complex; tooth disease.  
 OS Streptococcus mutans.  
 PN US2005031633-A1.  
 XX  
 PD 10-FEB-2005.  
 XX  
 PF 09-MAR-2004; 2004US-00797821.  
 XX  
 PR 13-APR-1998; 98US-0081550P.  
 PR 08-JAN-1999; 98US-0115142P.  
 PR 12-APR-1999; 99US-00290049.  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 PR 07-MAR-2003; 2003US-00383930.  
 XX (SMIT/) SMITH D J.  
 PA (TAUB/) TAUBMAN M A.  
 XX  
 PI Smith DJ, Taubman MA;  
 XX WPI; 2005-151644/16.  
 DR  
 XX  
 PT New composition comprising a fragment of a glucan binding protein-B  
 PT (GbpB) that binds to MHC class II protein, and a biocompatible  
 PT microparticle, useful for producing an antibody (claimed) for immunizing  
 PT mammals against dental caries.  
 XX  
 PS Claim 3; SEQ ID NO 30; 73pp; English.  
 CC The invention relates to a composition comprising a fragment of a glucan  
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
 CC fragment binds to a major histocompatibility complex (MHC) class II



CC protein. The composition is useful for producing an antibody for  
CC immunizing mammals against dental caries. This sequence corresponds to a  
CC Streptococcus mutans GbpB protein of the invention.  
XX  
SQ Sequence 431 AA;  
  
Query Match 98.7%; Score 2083.5; DB 9; Length 431;  
Best Local Similarity 99.1%; Pred. No. 3.6e-132;  
Matches 428; Conservative 2; Mismatches 1; Indels 1; Gaps 1;  
  
QY 1 MKKRILSAVLVSGVTLSATLSAIAKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60  
DB 1 MKKRILSAVLVSGVTLSATLSAIAKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60  
  
QY 61 QVSALQTQQAELQAEQNRLEAQSAITLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120  
DB 61 QVSALQTQQAELQAEQNRLEAQSAITLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120  
  
QY 121 INAIINSKVSDAINRVSAIREVVSANERKMLQOQEQKAAVEQKQENQAAINTVAANQE 180  
DB 121 INAIINSKVSDAINRVSAIREVVSANERKMLQOQEQKAAVEQKQENQAAINTVAANQE 180  
  
QY 181 TTAQNTNALNTQQAQLLEAQLNLQAEITTAQDOKATLVAQKAAEEAARQAARAAAE 240  
DB 181 TTAQNTNALNTQQAQLLEAQLNLQAEITTAQDOKATLVAQKAAEEAARQAARAAAE 240  
  
QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAADNTQAAQTGDSSTQSAQAQVNN 300  
DB 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAADNTQAAQTGDSSTQSAQAQVNN 300  
  
QY 301 SDOESTTATAQPSASASTAAVAANTSSANTYPAGCTGWGKSLAPWVGNVYWGNGGQWA 360  
DB 301 SDOESTTATAQPSASASTAAVAANTSSANTYPAGCTGWGKSLAPWVGNVYWGNGGQWA 360  
  
QY 361 ASAAAGYRGVSPSAGAVVNDGGYGVHVVYVTVGGGQIOVQEAANYAGNQSIGNYRGW 420  
DB 361 ASAAAGYRGVSPSAGAVVNDGGYGVHVVYVTVGGGQIOVQEAANYAGNQSIGNYRGW 420  
  
QY 421 FNPGSVSYIYPN 432  
DB 420 FNPGSVSYIYPN 431  
  
RESULT 7  
ADD93649  
ID ADD93649 standard; protein; 431 AA.  
XX  
AC ADD93649;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Streptococcus mutans glucan binding protein-B.  
XX  
KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.  
XX  
OS Streptococcus mutans.  
XX  
FH Key Location/Qualifiers  
FT Region 6..25 /note= "HLA-binding peptide"  
FT Region 16..35 /note= "HLA-binding peptide"  
FT Region 33..52 /note= "HLA-binding peptide"  
FT Region 37..56 /note= "HLA-binding peptide"  
FT Region 48..67 /note= "HLA-binding peptide"  
FT Region 52..71 /note= "HLA-binding peptide"  
FT Region 88..107 /note= "HLA-binding peptide"  
FT Region 113..132

FT Region /note= "HLA-binding peptide"  
FT 117..136  
FT Region /note= "HLA-binding peptide"  
FT 137..156  
FT Region /note= "HLA-binding peptide"  
FT 174..193  
FT Region /note= "HLA-binding peptide"  
FT 194..213  
FT Region /note= "HLA-binding peptide"  
FT 214..233  
FT Region /note= "HLA-binding peptide"  
FT 248..267  
FT Region /note= "HLA-binding peptide"  
FT 289..308  
FT Region /note= "HLA-binding peptide"  
FT 306..325  
FT Region /note= "HLA-binding peptide"  
FT 311..330  
FT Region /note= "HLA-binding peptide"  
FT 349..368  
FT Region /note= "HLA-binding peptide"  
FT 365..384  
FT Region /note= "HLA-binding peptide"  
FT 383..402  
FT Region /note= "HLA-binding peptide"  
FT 403..422  
FT Region /note= "HLA-binding peptide"  
FT  
FT  
XX  
XX  
PN W02003075845-A2.  
XX  
PD 18-SEP-2003.  
XX  
XX  
PF 07-MAR-2003; 2003WO-US006962.  
XX  
PR 07-MAR-2002; 2002US-0363209P.  
PR 08-AUG-2002; 2002US-0402483P.  
XX  
XX  
PA (FORS-) FORSYTH INST.  
XX  
XX  
PI Smith DJ, Taubman MA;  
XX  
XX  
DR WPI; 2003-845091/78.  
DR GENBANK; AY046410.  
XX  
XX  
PT Composition useful as vaccines for dental caries comprises a fragment of  
PT a glucan binding protein-B binding to a major histocompatibility complex  
PT class II protein.  
XX  
PS Claim 6; Page 7; 49pp; English.  
XX  
XX  
CC The present sequence is the protein sequence of the glucan binding  
CC protein-B (GbpB) of Streptococcus mutans strain SK32. The sequence  
CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
CC invention provides immunogenic compositions and vaccines for dental  
CC caries. The compositions comprise major histocompatibility complex (MHC)  
CC class II protein-binding GbpB peptides, especially HLA-binding peptides,  
CC covalently linked with peptide subunits of a glucosyltransferase. The  
CC compositions are used in a claimed method of eliciting production of an  
CC antibody in a mammal. Diepitopic or multiepitopic polypeptides can be  
CC prepared synthetically or by recombinant DNA technology. Antibodies  
CC raised against MHC class II binding fragments of GbpB can be used in  
CC passive immunisation.  
XX  
SQ Sequence 431 AA;  
  
Query Match 98.6%; Score 2079.5; DB 7; Length 431;  
Best Local Similarity 98.8%; Pred. No. 6.7e-132;  
Matches 427; Conservative 2; Mismatches 1; Indels 1; Gaps 1;  
  
QY 1 MKKRILSAVLVSGVTLSATLSAIAKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60  
DB 1 MKKRILSAVLVSGVTLSATLSAIAKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60

QY 61 QVSALQTQOAELEAQNORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSNAAATSY 120  
DB 61 QVSALQTQOAELEAQNORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSNAAATSY 120  
QY 121 INAIINSKSVSDAINRVSAIREVVSANERKMLQQEQEDKAAVEQKQOENQAATNTVAANOE 180  
DB 121 INAIINSKSVSDAINRVSAIREVVSANERKMLQQEQEDKAAVEQKQOENQAATNTVAANOE 180  
QY 181 TTAQNTNALNTQOAELEAQNORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSNAAATSY 240  
DB 181 TTAQNTNALNTQOAELEAQNORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSNAAATSY 240  
QY 241 KAAAEKALQEQAAQAAQVAA-NNNTQATDASDQAAAAADNTQAAOTGSDTDSQAAAAQAAVNN 300  
DB 241 KAAAEKALQEQAAQAAQVAA-NNNTQATDASDQAAAAADNTQAAOTGSDTDSQAAAAQAAVNN 299  
QY 301 SDOESTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVNTWNGGGQWA 360  
DB 301 SDOESTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVNTWNGGGQWA 359  
QY 361 ASAAAAGYRVGSTPSAGAVAVWVNDGGYGHVAVVTGVQGGQIQVQEAANYAGNQSIGNYRGW 420  
DB 361 ASAAAAGYRVGSTPSAGAVAVWVNDGGYGHVAVVTGVQGGQIQVQEAANYAGNQSIGNYRGW 419  
QY 421 FNPQSVSYIYPN 432  
DB 420 FNPQSVSYIYPN 431

## RESULT 8

AD337272

ID AD337272 standard; protein; 431 AA.

AC AD337272;

DT 21-APR-2005 (first entry)

DE Streptococcus mutant glucan binding protein B variant #1.

XX immunogenicity; immune stimulation; glucan binding protein-B;  
XX microparticle; major histocompatibility complex; tooth disease.

OS Streptococcus mutans.

XX US2005031633-A1.

PN 10-FEB-2005.

XX 09-MAR-2004; 2004US-00797821.

XX 13-APR-1998; 98US-0081550P.

PR 08-JAN-1999; 99US-0115142P.

PR 12-APR-1999; 99US-00290049.

PR 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

PR 07-MAR-2003; 2003US-00383930.

XX (SMIT/) SMITH D J.

PA (TAUB/) TAUBMAN M A.

PI Smith DJ, Taubman MA;

XX WPI; 2005-151644/16.

DR New composition comprising a fragment of a glucan binding protein-B  
XX (GbpB) that binds to MHC class II protein, and a biocompatible  
PT microparticle, useful for producing an antibody (claimed) for immunizing  
PT mammals against dental caries.

XX Claim 3; SEQ ID NO 29; 73pp; English.

PS The invention relates to a composition comprising a fragment of a glucan  
XX binding protein-B (GbpB) and a biocompatible microparticle, where the  
CC

CC fragment binds to a major histocompatibility complex (MHC) class II  
CC protein. The composition is useful for producing an antibody for  
CC immunizing mammals against dental caries. This sequence corresponds to a  
CC Streptococcus mutans GbpB protein of the invention.

XX Sequence 431 AA;

QY Query Match 98.6%; Score 2079.5; DB 9; Length 431;

DB Best Local Similarity 98.8%; Pred. No. 6.7e-132; Indels 1; Gaps 1;

QY Matches 427; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTATLSAIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIIG 60

DB 1 MKKRILSAVLVSGVTLSSTATLSAIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIIG 60

QY 61 QVSALQTQOAELEAQNORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSNAAATSY 120

DB 61 QVSALQTQOAELEAQNORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSNAAATSY 120

QY 121 INAIINSKSVSDAINRVSAIREVVSANERKMLQQEQEDKAAVEQKQOENQAATNTVAANOE 180

DB 121 INAIINSKSVSDAINRVSAIREVVSANERKMLQQEQEDKAAVEQKQOENQAATNTVAANOE 180

QY 181 TTAQNTNALNTQOAELEAQNORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSNAAATSY 240

DB 181 TTAQNTNALNTQOAELEAQNORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSNAAATSY 240

QY 241 KAAAEKALQEQAAQAAQVAA-NNNTQATDASDQAAAAADNTQAAOTGSDTDSQAAAAQAAVNN 300

DB 241 KAAAEKALQEQAAQAAQVAA-NNNTQATDASDQAAAAADNTQAAOTGSDTDSQAAAAQAAVNN 299

QY 301 SDOESTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVNTWNGGGQWA 360

DB 301 SDOESTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVNTWNGGGQWA 359

QY 361 ASAAAAGYRVGSTPSAGAVAVWVNDGGYGHVAVVTGVQGGQIQVQEAANYAGNQSIGNYRGW 420

DB 361 ASAAAAGYRVGSTPSAGAVAVWVNDGGYGHVAVVTGVQGGQIQVQEAANYAGNQSIGNYRGW 419

QY 421 FNPQSVSYIYPN 432

DB 420 FNPQSVSYIYPN 431

## RESULT 9

AEB91500

ID AEB91500 standard; protein; 431 AA.

XX AC AEB91500;

DT 20-OCT-2005 (first entry)

XX Microbial pathogen adhesin protein sequence, SEQ ID NO:210.

DE algorithm; adhesin; pharmaceutical; vaccine; drug screening;

XX bordetella pertussis infection; antibacterial; pneumonia;

KW antiinflammatory; respiratory-gen.; gastric ulcer; antiulcer;

KW gastrointestinal-gen.; urinary tract infection; antimicrobial; uropathic.

XX Streptococcus mutans.

XX WO2005076010-A2.

XX PD 18-AUG-2005.

XX 07-FEB-2005; 2005WO-IN000037.

XX 06-FEB-2004; 2004IN-DE000173.

PR 20-JUL-2004; 2004US-0589227P.

XX (COUL ) COUNCIL SCI &amp; IND RES SOUTH AFRICA.

PA Sachdeva G, Kumar K, Jain P, Brahmachari SK, Ramachandran S;

PI





CC The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins  
XX  
SQ

Query Match 58.5%; Score 1234.5; DB 5; Length 447;  
Best Local Similarity 59.7%; Pred. No. 6.7e-75;  
Matches 276; Conservative 43; Mismatches 98; Indels 45; Gaps 9;  
QY 1 MKKRILSAVLVSGVTLSATTLSAIAKADFDQAIASQDSKINNLTAAQQAQAQVNTIIG 60  
Db 1 MKKRILSAVLVSGVTLTAAV--TVNADDFSKIAATDSVINTLSQQAQAQVNTAIG 58  
QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQAQSAQKSNATSY 120  
Db 59 QVGALESQSELEAQAQLEAVSQQLGQIEIQTLSKNKIVARNESLKQVRSQAKGN-LTNY 117  
QY 121 INAINSKSVSDAIRVSAIRVSVANERKMLQOQEDKAAVEQKQOENQAINTVAANOE 180  
Db 118 INTILNSKSVSDAIRVSAIRVSVANERKMLQOQEDKAAVEQKQOENQAINTVAANKQ 177  
QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQDQKATLVQAQAAAEFAAQAAAQAABEA 240  
Db 178 ALENKAALATQRAQLEAAQLESLAQLTTVQNEKASLIQAKAQAEEAARKAQAQAABEA 237  
QY 241 KAAAEKALQEAQAQAQAQAANNNTQATDASDQQAADNTQAAQTGDSQSAQAQVNN 300  
Db 238 KAQAQAQAQESVAKAQAQA--QVESATAPTETVQTPTEIKPSNLTAT 285  
QY 301 SDQ---ESTTATA-----AQPSASSASTAA-----VAANTSSANTYPAG 336  
Db 286 SSATTVAITTTATNEPKVTQSVVTKAVEAPKAVVSTTPRAVSKPVVRSYDSNTYPNG 345  
QY 337 QCTWGVKSLAPVGNVWNGGQWAAASAAAAGYRVGSTPSAGAVAV--NDGGYGHVAVYT 394  
Db 346 QCTWGAKSMAVGNVWNGNAGASARAAGYSVGTTPRVGAVAVWPYDGGYGHVAVYT 405  
QY 395 GV-QGGQIQVQEAQVAGNOSIGNYRGWNP---GSVSYIYPN 432  
Db 406 SVANNSSIQVESNYAGNMSIGNYRGSFNPSAGSVYIYPN 447

RESULT 13  
ADU69524  
ID ADU69524 standard; protein; 447 AA.  
XX  
AC ADU69524;  
XX  
DT 10-FEB-2005 (first entry)  
XX  
DE S agalactiae hyperimmune serum reactive antigen seqid 219.  
XX  
KW immune stimulation; antigen; bacterial surface display;  
KW hyperimmune serum reactive antigen; vaccine; bacterial infection;  
KW antibacterial; infection.  
XX  
OS Streptococcus agalactiae.  
XX

PN W02004099242-A2.  
XX 18-NOV-2004.  
XX 06-MAY-2004; 2004WO-EP004856.  
XX 07-MAY-2003; 2003EP-00450112.  
PR 28-NOV-2003; 2003EP-00450266.  
XX (INTE-) INTERCELL AG.  
XX  
PI Meinke A, Nagy E, Hanner M, Horky M, Kallenda S, Prustomersky S;  
XX WPI; 2004-821662/81.  
DR N-PSDB; ADU69307.  
XX  
PT New nucleic acid molecule encoding a hyperimmune serum reactive antigen,  
PT useful for the manufacture of a vaccine against Streptococcus agalactiae  
PT infection.  
XX  
XX Claim 14; SEQ ID NO 219; 221pp; English.  
PS  
PS The invention describes an isolated nucleic acid molecule encoding a  
CC hyperimmune serum reactive antigen or its fragment. Also described are: a  
CC vector comprising the nucleic acid molecule; a host cell comprising the  
CC vector; a hyperimmune serum-reactive antigen comprising a sequence  
CC encoded by the nucleic acid molecule and consisting of e.g., 85, 299, 467  
CC or 812 amino acids; fragments of hyperimmune serum-reactive antigens  
CC consisting of peptides comprising e.g., 76, 134, 221 or 576 amino acids;  
CC a process for producing a Streptococcus agalactiae hyperimmune serum  
CC reactive antigen; a process for producing a cell that expresses a S.  
CC agalactiae hyperimmune serum reactive antigen; a pharmaceutical  
CC composition, especially a vaccine, comprising the hyperimmune serum-  
CC reactive antigen or nucleic acid molecule; an antibody that binds at  
CC least to a selective part of the hyperimmune serum-reactive antigen; a  
CC hybridoma cell line, which produces the antibody; a method for producing  
CC the antibody; an antagonist that binds to the hyperimmune serum-reactive  
CC antigen; a method for identifying an antagonist capable of binding to the  
CC hyperimmune serum-reactive antigen; a method for identifying an  
CC antagonist capable of reducing or inhibiting the interaction activity of  
CC a hyperimmune serum-reactive antigen to its interaction partner; a  
CC process for in vitro diagnosing a disease related to expression of the  
CC hyperimmune serum-reactive antigen; and a process for in vitro diagnosis  
CC of a bacterial infection, especially a S. agalactiae infection. The  
CC hyperimmune serum reactive antigen is useful for isolating, purifying  
CC and/or identifying an interaction partner of the hyperimmune serum  
CC reactive antigen. The hyperimmune serum reactive antigen is useful for  
CC generating a peptide binding to the hyperimmune serum reactive antigen,  
CC where the peptide comprises anticalines, or for the manufacture of a  
CC functional nucleic acid comprising aptamers or spiegelmers. The nucleic  
CC acid molecule is useful for the manufacture of a functional ribonucleic  
CC acid comprising ribozymes, antisense nucleic acids or siRNA. The nucleic  
CC acid molecule, hyperimmune serum-reactive antigen or antibody is useful  
CC for the manufacture of a vaccine against S. agalactiae infection. This is  
CC the amino acid sequence of a Streptococcus agalactiae hyperimmune serum  
CC reactive antigen.  
XX  
SQ

Sequence 447 AA;  
Query Match 58.5%; Score 1234.5; DB 8; Length 447;  
Best Local Similarity 59.7%; Pred. No. 6.7e-75;  
Matches 276; Conservative 43; Mismatches 98; Indels 45; Gaps 9;  
QY 1 MKKRILSAVLVSGVTLSATTLSAIAKADFDQAIASQDSKINNLTAAQQAQAQVNTIIG 60  
Db 1 MKKRILSAVLVSGVTLTAAV--TVNADDFSKIAATDSVINTLSQQAQAQVNTAIG 58  
QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQAQSAQKSNATSY 120  
Db 59 QVGALESQSELEAQAQLEAVSQQLGQIEIQTLSKNKIVARNESLKQVRSQAKGN-LTNY 117  
QY 121 INAINSKSVSDAIRVSAIRVSVANERKMLQOQEDKAAVEQKQOENQAINTVAANOE 180  
XX  
XX

Db 118 INTILNSKSVSDAVNRVVAIREVVSANEKMLAQEQADKAALEAKQIENQNAINTVAANKQ 177  
 QY 181 TTAQNTNALNTQQAQLEAAQLNLQALTTAQQOKATLVAQKAAABEAAARQAAAAQAAABA 240  
 Db 178 AIENKKAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQABEAAARKAAEAQAAABA 237  
 QY 241 KAAAEAKALQEQAAQAAQAAANNNNTQATDASDQQAADNTQAAOTGDSTDSQAQAQAVNN 300  
 Db 238 KAAAEAKAQESVAKAQAAA-----QVESATAPTETVQTPRTEIKPSNLTAT 285  
 QY 301 SDQ-----ESTTATA-----AOPSASSASTAA-----VAANTSSANTYPAG 336  
 Db 286 SSATTVAITTTATATNEPKVTQPSVVTKAVEAPKAVVSSTPRAVSKPVVRSYDSSNTYPMG 345  
 QY 337 QCTGWGKSLAPVWGNVWNGWGQAAASAAAAGYRVGSTPSAGAVW--NDGGYGHVAVYT 394  
 Db 346 QCTWGAKSMAWVGNVWGNWGNQWASARAAGYSGVGTTPRVGAVAVWVPYDGGGYGHVAVT 405  
 QY 395 GV-QGGQTOVQEQANYAGNQSIGNYRGWNP---GSVSYIYN 432  
 Db 406 SVANNSSIQVMESYAGNMSIGNYRGSPNPSAGSVYIYN 447

RESULT 14

ID ADV88392 standard; protein; 447 AA.  
 AC ADV88392;  
 DT 24-FEB-2005 (first entry)  
 DE Streptococcus agalactiae protein sequence, SEQ ID 786.  
 XX Antibacterial; Vaccine; bacterial infection.  
 XX Streptococcus agalactiae.  
 XX FR2824074-A1.  
 XX 31-OCT-2002.  
 XX 26-APR-2001; 2001PR-00005642.  
 XX 26-APR-2001; 2001PR-00005642.  
 XX (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Glaser P, Rusniok C, Chevallier F, Frangeul L, Lalioui L;  
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;  
 XX WPI; 2004-101891/11.

Genomic nucleotide sequences encoding polypeptides of Streptococcus agalactiae for the development of vaccines, diagnostic tools, DNA chips and identification of therapeutic targets.

Claim 6; SEQ ID NO 786; 2687pp; French.

The present invention relates to novel Streptococcus agalactiae nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II; ADV87746-ADV8950). The nucleotide sequences encode polypeptides of S. agalactiae involved in the synthesis of amino acids, cell membranes, intermediate (central) metabolism, energetic metabolism, fatty acid and phospholipid metabolism, nucleotide metabolism including purines, pyrimidines and/or nucleosides, regulatory functions, replication, transcription, translation, protein transport, adaptation to atypical conditions, sensitivity to medicines and/or analogues, functions related to transposons, biosynthesis of cofactors, prosthetic groups and transporters, cell membrane proteins and cellular machinery. (I) are useful for the detection and/or amplification of nucleic acids. CC Pharmaceutical composition comprising (I) or (II) are useful for treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is

CC equivalent for the present basic patent FR2824074A1. WO200292818A2 contains 6617 sequence whereas the present patent only contains 2344 sequences.  
 XX SQ Sequence 447 AA;  
 Query Match 58.5%; Score 1234.5; DB 8; Length 447;  
 Best Local Similarity 59.7%; Pred No. 6.7e-75;  
 Matches 276; Conservative 43; Mismatches 98; Indels 45; Gaps 9;  
 QY 1 MKKRILSAVLVSGVTLSAATLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTIAG 60  
 Db 1 MKKRILSAVLVSGVTLTGTAAV--TVNADDFDSKIAATDSVINTLSGQQAQAQVNTAIKG 58  
 QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQTLSSKIARNESLKOQARSQAOKSNAATSY 120  
 Db 59 QVGALESQSQSELEAQAQLEAVSQQLGQGIQTLNKNIVARNESLKKQVRSQAQKGN-LTNY 117  
 QY 121 INAIINSKSVSDAINRVSAIREVVSANEKMLQOQOQODKAAVFEKOQENQAAINTVAANOE 180  
 Db 118 INTILNSKSVSDAVNRVVAIREVVSANEKMLAQEQADKAALEAKQIENQNAINTVAANKQ 177  
 QY 181 TTAQNTNALNTQQAQLEAAQLNLQALTTAQQOKATLVAQKAAABEAAARQAAAAQAAABA 240  
 Db 178 AIENKKAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQABEAAARKAAEAQAAABA 237  
 QY 241 KAAAEAKALQEQAAQAAQAAANNNNTQATDASDQQAADNTQAAOTGDSTDSQAQAQAVNN 300  
 Db 238 KAAAEAKAQESVAKAQAAA-----QVESATAPTETVQTPRTEIKPSNLTAT 285  
 QY 301 SDQ-----ESTTATA-----AOPSASSASTAA-----VAANTSSANTYPAG 336  
 Db 286 SSATTVAITTTATATNEPKVTQPSVVTKAVEAPKAVVSSTPRAVSKPVVRSYDSSNTYPMG 345  
 QY 337 QCTGWGKSLAPVWGNVWNGWGQAAASAAAAGYRVGSTPSAGAVW--NDGGYGHVAVYT 394  
 Db 346 QCTWGAKSMAWVGNVWGNWGNQWASARAAGYSGVGTTPRVGAVAVWVPYDGGGYGHVAVT 405  
 QY 395 GV-QGGQTOVQEQANYAGNQSIGNYRGWNP---GSVSYIYN 432  
 Db 406 SVANNSSIQVMESYAGNMSIGNYRGSPNPSAGSVYIYN 447

RESULT 15

ADV81808  
 ID ADV81808 standard; protein; 447 AA.  
 AC ADV81808;  
 DT 24-FEB-2005 (first entry)  
 DE Streptococcus agalactiae protein, SEQ ID 2949.  
 XX Antibacterial; vaccine; bacterial infection.  
 XX Streptococcus agalactiae.  
 XX WO200292818-A2.  
 XX 21-NOV-2002.  
 XX 26-APR-2002; 2002WO-IB003059.  
 XX 26-APR-2001; 2001PR-00005642.  
 XX (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Glaser P, Rusniok C, Chevallier F, Frangeul L, Lalioui L;  
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
 XX WPI; 2004-101891/11.

Genomic nucleotide sequences encoding polypeptides of Streptococcus agalactiae for the development of vaccines, diagnostic tools, DNA chips and identification of therapeutic targets.

Claim 6; SEQ ID NO 2949; 439pp; French.

The present invention relates to novel Streptococcus agalactiae nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The nucleotide sequences encode polypeptides of S. agalactiae involved in the synthesis of amino acids, cell membranes, intermediate (central) metabolism, energetic metabolism, fatty acid and phospholipid metabolism, nucleotide metabolism including purines, pyrimidines and/or nucleosides, regulatory functions, replication, transcription, translation, protein transport, adaptation to atypical conditions, sensitivity to medicines and/or analogues, functions related to transposons, biosynthesis of cofactors, prosthetic groups and transporters, cell membrane proteins and cellular machinery. (I) are useful for the detection and/or amplification of nucleic acids. Pharmaceutical composition comprising (I) or (II) are useful for treatment of a bacterial S. agalactiae infection. The complete genome of Streptococcus agalactiae is given in ADV81204. Note: The present patent is an equivalent for the basic patent FR2824074A1, which contains only 2344 sequences.

Sequence 447 AA;

Query Match 58.5%; Score 1234.5; DB 8; Length 447;  
Best Local Similarity 59.7%; Pred. No. 6.7e-75;  
Matches 276; Conservative 43; Mismatches 98; Indels 45; Gaps 9;

1 MKKRILSAVLVSGVTLSSATTLSATKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQC 60  
1 MKKRILSAVLVSGVTLGTAAV--TVNADDFDSKIATDSVINTLSGQQAQAQNVTAIKG 58  
61 QVSALQTQCAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQOARSAOKSNAATSY 120  
59 QVGALESQSELEAQAQLEAQAQVNSQQLGQEIQTLLSNKIVARNESLKQVRSQKGN-LTNY 117  
121 INAIINSKVSDAINRVSAIREVVSANERKMLQEQEQDKAAVEKQEQENQAAINTVAANOE 180  
118 INTILNSKVSDAVNRVVAIREVVSANERKMLAQEQEADKALEAKQIENQNAINTVAANKQ 177  
181 TTAQNTNALNTQQAQLEAQAQLEAQLMLQELTTAQDOKATLVQAQAAAEAAQAAQAQAAEA 240  
178 ATEENKAALATQRAQLEAQAQLEAQLTTVQNEKASLIQAQAQEAQAAQAAEAQAAEA 237  
241 KAAAEAKALQEQAAQQAQAQAANNNTQATDASDQQAQAADNTQAAQTGSDTQDQAQAQVNN 300  
238 KAAQAEAKAQAESVAKAQAAA-----QVESATAPTETVQTQPTETKPSNLTAT 285  
301 SDQ---ESTTATA-----AQPSASSASTAA-----VAANTSSANTYPAG 336  
286 SSATTVAITTTATATNEPKVTQPSVVTKAVEAPKAVVSTPRAVSKPVRSYDSSNTYPNG 345  
337 QCTWGVKSLAPVWGNVWNGGQWAAASAAAAGYRVGSTPSAGAVVW--NDGGYGHVAVYT 394  
346 QCTWGAKSMAWVGNVWGNWGNWGAASARAAGYSVGTTPRVGAVVWVWPDGGYGHVAVYT 405  
395 GV-QGGQIQVQEAQYAGNOSIGNYRWENP---GSUSYIYPN 432  
406 SVANNSSIQVMESNAGNMSIGNYRGSFNPSPASGSVYIYPN 447

Search completed: February 15, 2006, 18:02:59  
Job time : 117.754 secs

**THIS PAGE BLANK (USPTO)**







|  |     |  |     |
|--|-----|--|-----|
| Db   | 1   | MKKKIISATLMSTVLSAAAPLSGVYADT-NSDIARQDATISSAQAKAQAQVDSLQS       | 59  |
| Qy   | 61  | QVSALOTQAELEAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAOKSNAATSY   | 120 |
| Db   | 60  | KVDSLQOKQASTYQAQIAKIESEKALNAQIATLINESIAERTKILEQAARSQAUNSSATNY  | 119 |
| Qy   | 121 | INAINIINSKVSDAINRVSAIREVVSANEKMLQQOEQDKAAVEQKQOENQAANTVAANOQE  | 180 |
| Db   | 120 | MDAVNKSLSLTDVIQKVTAIATVSSANKQMLEQOEKEQKLSQKSETVKKNYNQFVLSLQ    | 179 |
| Qy   | 181 | TTAONTNALNTQQAOLEAQAQLNLOAELTTAODOKATLVAQKAAAEAAARQAAAAA       | 240 |
| Db   | 180 | SLDSQAQELTSQAELKVATLNYQATYATQDKKOSLLDEKAAEAKAAQAKKQAAEYA       | 239 |
| Qy   | 241 | KAAAEAKALQEOAAQAQAANNNNTQATDASDQAAAAADNTQAAQTGSDTDOSSAQ--AV    | 298 |
| Db   | 240 | Q-----QKEAKAQAATAATTKAV---BEATSVSSSQASQSSSSNTSSNNTSS           | 288 |
| Qy   | 299 | NNSDQESTTATAAQPSASSASTAAVAANTSS-----ANTYPAG                    | 336 |
| Db   | 289 | NSSSSSNSSSSNSSSSNGGSTNTGNNAAGTNGTGGSSSGINSTPIANPYAGG           | 348 |
| Qy   | 337 | QCT---WG-----VKSLAPWGVNMGNGGWAASAAAAG--YRVGSTPAGAVAV--         | 381 |
| Db   | 349 | GCTDYVWQYFAAQGIYIRIMP-----GNGGQWATNGPAQGVHLHYVGAAP--GVIASSF    | 400 |
| Qy   | 382 | -----WNDGGYGHVAVYVTVG--QGQGIQVQEAANYAGNQSIGNVRGFWFGSVSYIYPN    | 432 |
| Db   | 401 | SADPVGYANSPIGHVAIVKSVNSNGTITKEGGY-GTTWGHGR-TVSASGVTFILMPN      | 456 |
| RESULT 5   |     |  |     |
| S05542   |     |  |     |
| hypothetical protein, 54K - Enterococcus faecium   |     |  |     |
| C:Species: Enterococcus faecium  |     |  |     |
| C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004              |     |  |     |
| C:Accession: S05542  |     |  |     |
| R:Fuerst, P.; Moesch, H.U.; Solioz, M.   |     |  |     |
| Nucleic Acids Res. 17, 6724, 1989  |     |  |     |
| A:Title: A protein of unusual composition from Enterococcus faecium.                     |     |  |     |
| A:Reference number: S05542; PMID:89385998; PMID:2780297                                  |     |  |     |
| A:Molecule type: DNA   |     |  |     |
| A:Accession: S05542  |     |  |     |
| A:Residues: 1-507 <FUE>  |     |  |     |
| A:Cross-references: UNIPROT:PI3692; UNIPARC:UPI000016F6FC; GB:X16421; EMBL:M26048; NID:9 |     |  |     |
| A>Note: the authors translated the codon CGT for residues 221 and 223 as Lys             |     |  |     |
| Query Match 25.8%; Score 543.5; DB 2; Length 507;  |     |  |     |
| Best Local Similarity 28.9%; Pred. No. 1.5e-19;  |     |  |     |
| Matches 147; Conservative 90; Mismatches 170; Indels 101; Gaps 12;                       |     |  |     |
| Qy   | 10  | LVSGVTLSSATLSAISKADDFDAQIASODSKINNLTAAQQAQAQVNTTQGVVSALOTQO    | 69  |
| Db   | 1   | MLSSIALTVAGSPIAAADDFSQIQQQDKKIADLQNOQAQSAQSQIEALEGGVSAINTKA    | 60  |
| Qy   | 70  | AELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAOKSNAATSYINAINIINSK | 129 |
| Db   | 61  | QDLTLTKQDTLRKESALQKEIKDLQERIEKREATIQKARETVQKNTSNYIDAVLNADS     | 120 |
| Qy   | 130 | VSDAINRVSAIREVVSANEKMLQQOEQDKAAVEQKQOENQAANTVAANOETIAQNTNAL    | 189 |
| Db   | 121 | LADAVGRIQAMSTIVKANQDLVQOQKEDQKQAEKKAENAKQKELADNQAALLESQKGLD    | 180 |
| Qy   | 190 | NTQQAOLEAQAQLNLOAELTTAODOKATLVAQKAAAEAAARQAAAAA                | 235 |
| Db   | 181 | LAKQADNLNLTSLAAEQATAEKDKADLNKKAEEAEQARIREQARLAEARQQAQAEK       | 240 |
| Qy   | 236 | AAAEKAAAEAKALQEOAAQAQAANNNNTQATDASDQAAA-----AAEEAARQAAAAQ      | 282 |
| Db   | 241 | AEKEAREQAAQAQATQALSSASTTTESSSAQSSSEESKAPESSTTESTESTSTTTE       | 300 |
| Qy   | 283 | AAQTGSDTDOSSA----AQAVNNSDQESTTATAAQPSASSASTAAVAANTSANT-----    | 332 |

|  |     |   |     |
|--|-----|---|-----|
| Db   | 301 | NSSTGSSSTESSSTBEESTVPESSTQESTPANTSESSSSSNTNVNNNTNNSTNNSTT     | 360 |
| Qy   | 333 | -----YPAGQCTWGVKSLA-----PWVGNVWNG-----                        | 356 |
| Db   | 361 | NNNNNNNTVTPAPTPTTPAPAPAPNPSGVSNGAAIVAEAYKYIGTPLYV---WGGKDPG   | 417 |
| Qy   | 357 | -----GOWAASAAAAAGYRVG-STPSAGAVAVVND--GGYGHVAVYVVG             | 395 |
| Db   | 418 | FDCSGFTRVYVLQVTRDIGMTVPQESAGTKISVSOAKAGDLLFWGSAGGTVHVAISLG    | 477 |
| Qy   | 396 | VOGGQIQVQEAANYAGNQSIGNVRGWFNP                                 | 423 |
| Db   | 478 | --GGQ-YIHAPQPGENVKVGVSQ-WYTP                                  | 501 |
| RESULT 6   |     |   |     |
| AC1763   |     |   |     |
| peptidoglycan lytic protein P45 [imported] - Listeria innocua (strain Clip11262)   |     |   |     |
| C:Species: Listeria innocua  |     |   |     |
| C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004  |     |   |     |
| C:Accession: AC1763  |     |   |     |
| R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  |     |   |     |
| Science 294, 849-852, 2001   |     |   |     |
| A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, C. |     |   |     |
| A:Title: Comparative genomics of Listeria species  |     |   |     |
| A:Reference number: AB1077; PMID:11679669  |     |   |     |
| A:Accession: AC1763  |     |   |     |
| A:Status: preliminary  |     |   |     |
| A:Molecule type: DNA   |     |   |     |
| A:Residues: 1-398 <GLA>  |     |   |     |
| A:Cross-references: UNIPROT:Q927Y8; UNIPARC:UPI00000CC935; GB:AL592022; PIDN:CAC97875.1;   |     |   |     |
| A:Experimental source: strain Clip11262  |     |   |     |
| C:Genetics:  |     |   |     |
| A:Gene: spl  |     |   |     |
| Query Match 17.3%; Score 364; DB 2; Length 398;  |     |   |     |
| Best Local Similarity 27.4%; Pred. No. 5.4e-11;  |     |   |     |
| Matches 115; Conservative 88; Mismatches 145; Indels 72; Gaps 16;  |     |   |     |
| Qy   | 1   | MKKRILSAV-LVSGVTLSSATLSAISKADDFDAQIASODSKINNLTAAQQAQAQVNTTIO  | 59  |
| Db   | 1   | MKKNTFIAISLAAVISLTPAFTTNVFA--DVNTDIQNDKINDIKSKTKGLQSLSLV      | 58  |
| Qy   | 60  | GOVSALOTQQAOLEAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAOKSNAATS | 119 |
| Db   | 59  | ADLEKAQEKAKSLOGEPDQTGKELQNLQDIKDINERIKERETVLKERARAWQKTSNSNA   | 118 |
| Qy   | 120 | YINAIINSKVSDAINRVSAIREVVSANEKMLQQOEQD-----KAAVEQKQOENQAAI     | 172 |
| Db   | 119 | YLEVILDAENLSDLVGRVSAVNNQLVDSKSILEDQKDEKALKTKQTAVKKQEEQATAI    | 178 |
| Qy   | 173 | NTVAANOETIAQNTNALNTQQAOLEAQAQLNLOAELTTAODOKATLVAQKAAAEAAARQA  | 232 |
| Db   | 179 | HEFEAAQ-----NKIEAQAKAEKAIVAQLAADQAENAKGLVSE--DKAAKE--         | 226 |
| Qy   | 233 | AAQAAAEKAAAEKALQEOAAQAQAANNNNTQATDASDQAAAAADNTQAAQTGSDTQ      | 292 |
| Db   | 227 | -----ATARATALRE-----ATDANVGQOTT--NTNA--SSNSKTS                | 258 |
| Qy   | 293 | SAAQAVNNSDQESTTATAAQPSASSASTAAVAANTSANTYPAG-----OCTWGVKSL     | 345 |
| Db   | 259 | NKVESTNNSEAP-----HATPSGGGYSAMIAARAQLGKPYLSLGATGPSAFPDCS-GFTSY | 313 |
| Qy   | 346 | ---APWVGNVWNGGQWAAASAAAAGYRVGSTPSAGAVVNM-DGGYGHVAVYVVGQGGI    | 401 |
| Db   | 314 | AFRAAGVSLPRTSGGOYAAAASKIS----ASQAKPGDLVPFNFGGGIAHVGIYVG--GGQM | 367 |
| RESULT 7   |     |   |     |









**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 17:57:21 ; Search time 123.762 Seconds  
(without alignments)  
2462.693 Million cell updates/sec

Title: US-10-797-821-31  
Perfect score: 2110  
Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNVRGMFPGSVIYIPN 432

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_05.80.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result # | No.    | Score | Query Match | Length | DB     | ID    | Description        |
|----------|--------|-------|-------------|--------|--------|-------|--------------------|
| 1        | 2110   | 100.0 | 432         | 2      | Q938V2 | STRMU | Q938V2 streptococc |
| 2        | 2099   | 99.5  | 432         | 2      | Q938V1 | STRMU | Q938V1 streptococc |
| 3        | 2083.5 | 98.7  | 431         | 2      | Q9AG98 | STRMU | Q9AG98 streptococc |
| 4        | 2079.5 | 98.6  | 431         | 2      | Q938V3 | STRMU | Q938V3 streptococc |
| 5        | 2079.5 | 98.6  | 431         | 2      | Q8DM3  | STRMU | Q8DM3 streptococc  |
| 6        | 2075.5 | 98.4  | 431         | 2      | Q938V0 | STRMU | Q938V0 streptococc |
| 7        | 1234.5 | 58.5  | 447         | 2      | Q9AKA4 | STRAG | Q9AKA4 streptococc |
| 8        | 1234.5 | 58.5  | 447         | 2      | Q8E2H1 | STRAS | Q8E2H1 streptococc |
| 9        | 1234.5 | 58.5  | 447         | 2      | Q8E7X9 | STRAS | Q8E7X9 streptococc |
| 10       | 1093   | 51.8  | 398         | 2      | Q5XEL1 | STRP6 | Q5XEL1 streptococc |
| 11       | 1093   | 51.8  | 398         | 2      | Q7CNO7 | STRP8 | Q7CNO7 streptococc |
| 12       | 1093   | 51.8  | 398         | 2      | Q8P316 | STRP3 | Q8P316 streptococc |
| 13       | 1092.5 | 51.8  | 474         | 2      | Q5M6K4 | STRT2 | Q5M6K4 streptococc |
| 14       | 1092   | 51.8  | 485         | 2      | Q5M212 | STRT1 | Q5M212 streptococc |
| 15       | 1087   | 51.5  | 398         | 2      | Q9A1Z8 | STRPY | Q9A1Z8 streptococc |
| 16       | 1059   | 50.2  | 392         | 2      | Q8DMY4 | STRR6 | Q8DMY4 streptococc |
| 17       | 1059   | 50.2  | 392         | 2      | Q97N55 | STRPN | Q97N55 streptococc |
| 18       | 946    | 44.8  | 211         | 2      | Q9ZAS7 | STRMU | Q9ZAS7 streptococc |
| 19       | 676.5  | 32.1  | 461         | 1      | USP45  | IACLC | P22865 lactococcus |
| 20       | 660    | 31.3  | 456         | 2      | Q9CDJ1 | LACIA | Q9CDJ1 lactococcus |
| 21       | 588.5  | 27.9  | 524         | 2      | Q9K2J9 | ENTFC | Q9K2J9 enterococcu |
| 22       | 575.5  | 27.3  | 516         | 1      | P54    | ENTFC | P13692 enterococcu |
| 23       | 549.5  | 26.0  | 576         | 2      | Q9KJ3  | ENTHR | Q9KJ3 enterococcu  |
| 24       | 549    | 26.0  | 482         | 2      | Q5M5M6 | STRT2 | Q5M5M6 streptococc |
| 25       | 544    | 25.8  | 470         | 2      | Q8QEB3 | 9LACT | Q8QEB3 lactococcus |
| 26       | 532    | 25.2  | 482         | 2      | Q93LK4 | ENTFA | Q93LK4 enterococcu |
| 27       | 511.5  | 24.2  | 461         | 2      | Q56SA7 | STRTR | Q56SA7 streptococc |
| 28       | 504.5  | 23.9  | 449         | 2      | Q93LK3 | ENTFA | Q93LK3 enterococcu |
| 29       | 436.5  | 20.7  | 211         | 2      | Q8DVU8 | STRMU | Q8DVU8 streptococc |
| 30       | 416    | 19.7  | 226         | 2      | Q93RG6 | STRIT | Q93RG6 streptococc |
| 31       | 416    | 19.7  | 544         | 2      | Q840W6 | STRMU | Q840W6 streptococc |

|  |   |   |      |                      |     |               |    |     |  |
|--|---|---|------|----------------------|-----|---------------|----|-----|--|
| RESULT 1   |   |   |      |                      |     |               |    |     |  |
| Q938V2_STRMU   |   |   |      |                      |     |               |    |     |  |
| ID   | Q938V2_STRMU  | PRELIMINARY;                            | PRT; | 432                  | AA. |               |    |     |  |
| AC   | Q938V2;   |   |      |                      |     |               |    |     |  |
| DT   | 01-DEC-2001   | (TrEMBLrel. 19, Created)                |      |                      |     |               |    |     |  |
| DT   | 01-DEC-2001   | (TrEMBLrel. 19, Last sequence update)   |      |                      |     |               |    |     |  |
| DT   | 01-MAR-2004   | (TrEMBLrel. 26, Last annotation update) |      |                      |     |               |    |     |  |
| DE   | Glucan-binding protein B.   |   |      |                      |     |               |    |     |  |
| OS   | Streptococcus mutans.   |   |      |                      |     |               |    |     |  |
| OC   | Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;              |   |      |                      |     |               |    |     |  |
| OC   | Streptococcus.  |   |      |                      |     |               |    |     |  |
| OX   | NCBI_TaxID=1309;  |   |      |                      |     |               |    |     |  |
| RN   | [1]   |   |      |                      |     |               |    |     |  |
| RP   | NUCLEOTIDE SEQUENCE.  |   |      |                      |     |               |    |     |  |
| RC   | STRAIN=15JP2;   |   |      |                      |     |               |    |     |  |
| RA   | Jin S., Duncan M.J., Taubman M.A., Smith D.J.;                        |   |      |                      |     |               |    |     |  |
| RT   | "Cloning of the gbpB gene from Streptococcus mutans.";                |   |      |                      |     |               |    |     |  |
| RL   | J. Dent. Res. 79:224-224(2000).                                       |   |      |                      |     |               |    |     |  |
| RN   | [2]   |   |      |                      |     |               |    |     |  |
| RP   | NUCLEOTIDE SEQUENCE.  |   |      |                      |     |               |    |     |  |
| RC   | STRAIN=15JP2;   |   |      |                      |     |               |    |     |  |
| RX   | MEDLINE=21481971; PubMed=11598068;                                    |   |      |                      |     |               |    |     |  |
| RX   | DOI=10.1128/IAI.69.11.6931-6941.2001;                                 |   |      |                      |     |               |    |     |  |
| RA   | Mattos-Grauer R.O., Jin S., King W.F., Chen T., Smith D.J.,           |   |      |                      |     |               |    |     |  |
| RA   | Duncan M.J.;  |   |      |                      |     |               |    |     |  |
| RT   | "Cloning of the Streptococcus mutans gene encoding glucan binding     |   |      |                      |     |               |    |     |  |
| RT   | protein B and analysis of genetic diversity and protein production in |   |      |                      |     |               |    |     |  |
| RT   | clinical isolates.";  |   |      |                      |     |               |    |     |  |
| RL   | Infect. Immun. 69:6931-6941(2001).                                    |   |      |                      |     |               |    |     |  |
| DR   | EMBL; AY046412; AAK94502.1; -; Genomic_DNA.                           |   |      |                      |     |               |    |     |  |
| DR   | InterPro; IPR007921; CHAP.  |   |      |                      |     |               |    |     |  |
| DR   | InterPro; IPR009148; Siba.  |   |      |                      |     |               |    |     |  |
| DR   | Pfam; PF05257; CHAP; 1.   |   |      |                      |     |               |    |     |  |
| DR   | PRINTS; PR01852; SIBAPROTEIN.   |   |      |                      |     |               |    |     |  |
| DR   | PROSITE; PS50911; CHAP; 1.  |   |      |                      |     |               |    |     |  |
| SQ   | SEQUENCE 432 AA; 44648 MW; E769B2504AE50E9 CRC64;                     |   |      |                      |     |               |    |     |  |
| Query Match 100.0%; Score 2110; DB 2; Length 432;            |   |   |      |                      |     |               |    |     |  |
| Best Local Similarity 100.0%; Pred. No. 2.7e-86;             |   |   |      |                      |     |               |    |     |  |
| Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |   |   |      |                      |     |               |    |     |  |
| QY   | 1   | MKKRILSAVLVSGVTLSAT                     | TL   | SAIKADDFDAQIASQDSKIN | LN  | TAQQAQAQVNTI  | Q  | 60  |  |
| Db   | 1   | MKKRILSAVLVSGVTLSAT                     | TL   | SAIKADDFDAQIASQDSKIN | LN  | TAQQAQAQVNTI  | Q  | 60  |  |
| QY   | 61  | QVSALQTOAEIQAENRLE                      | QA   | QSATLGGQIQTLSSKIVARN | ESL | KQARSQAKSNAAT | S  | 120 |  |
| Db   | 61  | QVSALQTOAEIQAENRLE                      | QA   | QSATLGGQIQTLSSKIVARN | ESL | KQARSQAKSNAAT | S  | 120 |  |
| QY   | 121   | INAIINSKSVSDAINRV                       | SAIR | EVVSAKMLQQQEQDQKAA   | VE  | QKQENQAINTVA  | AN | 180 |  |
| Db   | 121   | INAIINSKSVSDAINRV                       | SAIR | EVVSAKMLQQQEQDQKAA   | VE  | QKQENQAINTVA  | AN | 180 |  |
| QY   | 181   | TTAQNTNALNTQQAOL                        | EA   | AAQLNLQALTTAQDQKAT   | LV  | AQKAAAEAAQAAA | A  | 240 |  |

Q840x3 streptococc  
Q840v8 streptococc  
Q8dur7 streptococc  
Q5ml37 streptococc  
Q8e3f4 streptococc  
Q8dx4 streptococc  
Q92y78 listeria in  
Q71w83 listeria mo  
Q9re04 listeria mo  
Q81hv2 bacillus ce  
Q6m552 corynebacte  
Q73dg0 bacillus ce  
P40767 bacillus su  
Q8nqa0 corynebacte

```

Db 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAQQKATLVAKAAAEAAEAARQAAAAQAAEA 240
QY 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSQAAQAVNN 300
Db 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSQAAQAVNN 300
QY 301 SDOESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGOWA 360
Db 301 SDOESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGOWA 360
QY 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGQGGQIQVQAEYANTAGNQSIGNYRGW 420
Db 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGQGGQIQVQAEYANTAGNQSIGNYRGW 420
QY 421 FNPQSVSYIYPN 432
Db 421 FNPQSVSYIYPN 432

RESULT 2
Q938V1_STRMU
ID Q938V1_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V1;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RX [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/JAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL: AV046413; AAK94503.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS09111; CHAP; 1.
SQ SEQUENCE 432 AA; 44652 MW; 3F88ECB9A1F3BE4F CRC64;

Query Match 99.5%; Score 2099; DB 2; Length 432;
Best Local Similarity 99.3%; Pred. No. 8.3e-86;
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKRLTSLAVLGVGVTLSATLSATKADDFDAQIASQDSKINNLTAQQAQAAQVNTIQG 60
Db 1 MKKRLTSLAVLGVGVTLSATLSATLSAVKADDFDAQIASQDSKINNLTAQQAQAAQVNTIQG 60

QY 61 QVSALQTOQAEIQAENORLEAQSATLGGQIQITLSSKIVARNESLKQARSQAQSAATSY 120
Db 61 QVSALQTOQAEIQAENORLEAQSATLGGQIQITLSSKIVARNESLKQARSQAQSAATSY 120

QY 121 INAIINSSVSDAINRVSAIRVSVANEKMLQQQODKAAVEQKQENQAINTVAANOQ 180
Db 121 INAIINSSVSDAINRVSAIRVSVANEKMLHQEQDQKAAVEQKQENQAINTVAANOQ 180

```

## RESULT 3

```

Q9AG98_STRMU
ID Q9AG98_STRMU PRELIMINARY; PRT; 431 AA.
AC Q9AG98;
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-FEB-2005 (TREMELrel. 29, Last annotation update)
DE Immunodominant glycoprotein IDG-60 (Glucan-binding protein B).
GN Name=saga;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RX [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=21153617; PubMed=11254612;
RX DOI=10.1128/JAI.69.4.2493-2501.2001;
RA Chia J.S., Lee Y.Y., Huang P.T., Chen J.Y.;
RT "Identification of stress-responsive genes in Streptococcus mutans by
RT differential display reverse transcription-PCR.";
RN Infect. Immun. 69:2493-2501(2001).
RX [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=21481977; PubMed=11598074;
RX DOI=10.1128/JAI.69.11.6987-6998.2001;
RA Chia J.S., Chang L.Y., Shun C.T., Chang Y.Y., Chen J.Y.;
RT "A 60-kilodalton immunodominant glycoprotein is essential for cell
RT wall integrity and the maintenance of cell shape in Streptococcus
RT mutans.";
RN Infect. Immun. 69:6987-6998(2001).
RX [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RA Chia J.-S., Chang L.-Y., Lee Y.-Y., Chen J.-Y.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3VF4;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3VF4;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/JAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in

```



```

RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).
DR EMBL; AE014855; RAN57811.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 431 AA; 44620 MW; 2D1CA685248CCD3E CRC64;

Query Match      98.6%; Score 2079.5; DB 2; Length 431;
Best Local Similarity 98.8%; Pred. No. 6.1e-85;
Matches 427; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSATLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
DB 1 MKKRILSAVLVSGVTLSATLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60

QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQQAARSQKSNAAATSY 120
DB 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQQAARSQKSNAAATSY 120

QY 121 INAIINSKSVSDAINRVSAIREVVSANERKMLQQEQDQKAAVEQKQENQAATNTVAANQE 180
DB 121 INAIINSKSVSDAINRVSAIREVVSANERKMLQQEQDQKAAVEQKQENQAATNTVAANQE 180

QY 181 TIAQNTNALNTQQAQLEAAQNLQAEITTAQDQKATLVAQKAAAEAEARQAAAAQAAEA 240
DB 181 TIAQNTNALNTQQAQLEAAQNLQAEITTAQDQKATLVAQKAAAEAEARQAAAAQAAEA 240

QY 241 KAAAEKAKLQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGSDTQSAQAQAVNN 300
DB 241 KAAAEKAKLQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGSDTQSAQAQAVNN 300

QY 301 SDQESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 360
DB 301 SDQESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 360

QY 361 ASAAAAGRVGTPSAGAVAVWVNDGGYGHVAVTVTGQGGQIQVQENYAGNOSIGNYRGW 420
DB 361 ASAAAAGRVGTPSAGAVAVWVNDGGYGHVAVTVTGQGGQIQVQENYAGNOSIGNYRGW 420

QY 421 FNPGSVSYIYPN 432
DB 420 FNPGSVSYIYPN 431

RESULT 6
Q938V0_STRMU PRELIMINARY; PRT; 431 AA.
AC Q938V0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5SM3;
RC Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5SM3;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/JAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding

RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941 (2001).
DR EMBL; AY046414; AAK94504.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 431 AA; 44650 MW; 05D38D8DBC4609F CRC64;

Query Match      98.4%; Score 2075.5; DB 2; Length 431;
Best Local Similarity 98.6%; Pred. No. 9.2e-85;
Matches 426; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSATLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
DB 1 MKKRILSAVLVSGVTLSATLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60

QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQQAARSQKSNAAATSY 120
DB 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQQAARSQKSNAAATSY 120

QY 121 INAIINSKSVSDAINRVSAIREVVSANERKMLQQEQDQKAAVEQKQENQAATNTVAANQE 180
DB 121 INAIINSKSVSDAINRVSAIREVVSANERKMLQQEQDQKAAVEQKQENQAATNTVAANQE 180

QY 181 TIAQNTNALNTQQAQLEAAQNLQAEITTAQDQKATLVAQKAAAEAEARQAAAAQAAEA 240
DB 181 TIAQNTNALNTQQAQLEAAQNLQAEITTAQDQKATLVAQKAAAEAEARQAAAAQAAEA 240

QY 241 KAAAEKAKLQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGSDTQSAQAQAVNN 300
DB 241 KAAAEKAKLQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGSDTQSAQAQAVNN 300

QY 301 SDQESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 360
DB 301 SDQESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 360

QY 361 ASAAAAGRVGTPSAGAVAVWVNDGGYGHVAVTVTGQGGQIQVQENYAGNOSIGNYRGW 420
DB 361 ASAAAAGRVGTPSAGAVAVWVNDGGYGHVAVTVTGQGGQIQVQENYAGNOSIGNYRGW 420

QY 421 FNPGSVSYIYPN 432
DB 420 FNPGSVSYIYPN 431

RESULT 7
Q9AKA4_STRAG PRELIMINARY; PRT; 447 AA.
AC Q9AKA4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PcsB protein precursor.
GN Name=pcsB;
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=6313;
RX MEDLINE=21101799; PubMed=11157929;
RX DOI=10.1128/JB.183.4.1175-1183.2001;
RA Reinscheid D.J., Gottschalk B., Schubert A., Eikmanns B.J.,
RA Chhatwal G.S.;
RT "Identification and molecular analysis of PcsB, a protein required for
RT cell wall separation of group B streptococcus.";
RL J. Bacteriol. 183:1175-1183 (2001).
DR EMBL; AJ277292; CAC28144.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.

```

```
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Signal.
FT SIGNAL. 1 25 Potential.
FT CHAIN 26 447 PcsB protein.
SQ SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 58.5%; Score 1234.5; DB 2; Length 447;
Best Local Similarity 59.7%; Pred. No. 2.1e-47;
Matches 276; Conservative 43; Mismatches 98; Indels 45; Gaps 9;

QY 1 MKKRILSAVLVSGVTLSSTATLSAISKADDFDAQIASQDSKINNLTAAQQAQAQVNTIIG 60
DB 1 MKKRILSAVLVSGVTILGTAAV--TVNADDFDSKIATDSVINTLSGQAAAQNVTAIKG 58
QY 61 QVSALQTOQAELEAENORLEAQSATLGQOIOTLSKIVARNESLKQOASAKSNAATSY 120
DB 59 QVGALESQSELEAQAQLEAVSQQLGQEIOTLSNKIVARNESLKQVRSQKGN-LTNY 117
QY 121 INAINSKSVDAINRVSAIREVVSANEXMLQOEODKAAVEQKQOENQAANTVAANO 180
DB 118 INTILNSKSVDAVNVVAIREVVSANEXMLQOEADKAALEAKQIENQNAINTVAANKQ 177
QY 181 TTAQNTNALNTQOALEAQAQLMLQAEELTTAQDQKATLVAKAAAEBAARQAQAQA 240
DB 178 ALENKAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQAEBAAKAAEAQA 237
QY 241 KAAAEKALQEAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 300
DB 238 KAAEAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 285
QY 301 SDQ---ESTTATA-----AQSASASTAA-----VAANTSANTYPAG 336
DB 286 SSATTVAITTTATATNEPKVTQTSVVTKAVEAPKAVVSTTPRAVSKPVRSYSSNTYPMG 345
QY 337 QCTWGVKSLAPVWGNVGNWGOWAASAAAAGYRVGSTPSAGAVAVM--NDGGYGHVAVYT 394
DB 346 QCTWGAKSMAWVGNVGNWGNWGNWGNWGNWGNWGNWGNWGNWGNWGNWGNWGNW 405
QY 395 GV-QGGQIQVQAEANYAGNQSIGNYRGWFPN----GSVSYIYPN 432
DB 406 SVANNSSIQVMESNYAGNMSIGNYRGSNFSPASGSVYIYPN 447

RESULT 9
Q8E7X9 STRA3 PRELIMINARY; PRT; 447 AA.
AC Q8E7X9_
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE PcsB protein.
GN Name=pcsB; OrderedLocustNames=gbs0016;
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rusnik C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couvre E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513 (2002).
DR EMBL; AL766843; CAD45661.1; -; Genomic_DNA.
```

```
DR SagaList; qbs0016; -.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match      58.5%; Score 1234.5; DB 2; Length 447;
Best Local Similarity 59.7%; Pred. No. 2.1e-47;
Matches 276; Conservative 43; Mismatches 98; Indels 45; Gaps 9;

QY 1 MKKRILSAVLVSGVTLSSATTLISAIKADDFDAQIASQDSKINNLTAAQOAAQAVNTTIG 60
DB 1 MKKRILSAVLVSGVTLGTAAT--TVNADDFDKIAATDSVINTLGGQAAAQNVTAIRG 58

QY 61 QVSALOTOAAEQAEORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQAOKSNAATSY 120
DB 59 QVGALESQOSELEAQAQLEAVSQOQGQEIQTLSNKIVARNESLKKQVRSQAQGN-LTNY 117

QY 121 INAIINSKSVSDAINRVSAIREVVSANERKMLQQEQDQKAAVEQKQOENQAATNTVAANO 180
DB 118 INTILNSKSVSDAVNKNVVAIREVVSANERKMLAQEQADKAALEAKQIENQNAINTVVAANKQ 177

QY 181 TIAQNTNALNTQOALEAAQLEAQLNLOAELTTAQDQKATLVAQKAAAEARQAAAAQAAEA 240
DB 178 AIENKKAALATQRAQLEAAQLELSAQLTTVQNEKASLITQAKAQAEEAAKAAEAQAAEA 237

QY 241 KAAAEAKALOEAQAAQAAQAAANNNTQATDASDQQAADNTQAAOTGSDTQSAQAQAVNN 300
DB 238 KAQAQAKAQAQSAQAQAA-----QVESATAPETVTQPTETIKPSNLTAT 285

QY 301 SDQ-----ESTTATA-----AQPSASASTAA-----VAANTSSANTYPAG 336
DB 286 SSATTVATTATATNEPKVTQPSVVTKAVEAPKAVVSTPRAPKVPVRSYDSSNTYPMG 345

QY 337 QCTWGVKSLAPVGVNGWGGWAASAAAGYRGVSTPSAGAVAVW--NDGGYGHVAVVT 394
DB 346 QCTWGAQSMASVGVNGWGNQWASARAAGYSVGTTPRVGAVAVWVPYDGGYGHVAVVT 405

QY 395 GV-QGGQIQVQOQYANVAGNOSIGNYRGWFPN---GSVSVIYPN 432
DB 406 SVANNSSIQVNESYAGNWSIGNYRGSNFNPASGSGSVIYPN 447

RESULT 10
QXELL1_STRP6
ID QXELL1_STRP6 PRELIMINARY; PRT; 398 AA.
AC QXELL1_
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Secreted protein.
GN OrderedLocusNames=M6_Spy0017;
OS Streptococcus pyogenes (serotype M6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301450;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN-MGAS10394;
RX PubMed=15272401; DOI=10.1086/422697;
RA Banks D.J., Porcella S.F., Barbican K.D., Beres S.B., Phillips L.E.,
RA Voyich J.M., DeLeo F.R., Martin J.M., Somerville G.A., Musser J.M.;
RT "Progress toward characterization of the group A Streptococcus
RT metagenome: complete genome sequence of a macrolide-resistant serotype
RT M6 strain.";
RL J. Infect. Dis. 190:727-738 (2004).
DR EMBL; CP000003; AAT86152.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBAPROTEIN.
DR Pfam; PF05257; CHAP; 1.

DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match      51.8%; Score 1093; DB 2; Length 398;
Best Local Similarity 53.8%; Pred. No. 3.5e-41;
Matches 233; Conservative 58; Mismatches 106; Indels 36; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSSATTLISAIKADDFDAQIASQDSKINNLTAAQOAAQAVNTTIG 60
DB 1 MKKRILSAVLVSGVTLGTAAT--VGAEDLSTKIAKQDSIIISNLTEQKAAQNVSAQOA 57

QY 61 QVSALOTOAAEQAEORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQAOKSNAATSY 120
DB 58 QVSSIQSSQDQLKLTARNTLEALSRRFEIEIKALTSQIVARNELKNQARSATYNNETSQY 117

QY 121 INAIINSKSVSDAINRVSAIREVVSANERKMLQQEQDQKAAVEQKQOENQAATNTVAANO 180
DB 118 INALINSKSISDVNRLVAIRAVSANAKLLEQQKADKVSLEEKQAANGTAINTIAANWA 177

QY 181 TIAQNTNALNTQOALEAAQLEAQLNLOAELTTAQDQKATLVAQKAAAEARQAAAAQAAEA 240
DB 178 MAEENQNTLRTQANLEAATANLALQLASATEDKANLVAQKEAAEKAAAEALAQEQAAKV 237

QY 241 KAAAEAKALOEAQAAQAAQAAANNNTQATDASDQQAADNTQAAOTGSDTQSAQAQAVNN 300
DB 238 KA-----QSOQA-----QQAASVEAAKSAITPAQATPAQAQSSWA 272

QY 301 SDQESTTATAAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPVGVNGWGGWA 360
DB 273 IEPALTAQAA-PSARPOT-----SYDSSNTYPVQCTWGAQSLAPWAGNWNWGGWA 325

QY 361 ASAAAAGYRGVSTPSAGAVAVWVNDGGYGHVAVVTGQV--GGQIQVQOENYAGNOSTGNVRG 419
DB 326 YSAQAAGYRGVSTPMVGAIAVWVNDGGYGHVAVVVEVQSASSIRVMESYSGRYADHRG 385

QY 420 WFNPGSVSVIYPN 432
DB 386 WFNPTGVTFIYPH 398

RESULT 11
QXCNQ7_STRP8
ID QXCNQ7_STRP8 PRELIMINARY; PRT; 398 AA.
AC QXCNQ7_
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Putative secreted protein.
GN OrderedLocusNames=spyM18_0020;
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301451;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN-MGAS18232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108; DOI=10.1073/pnas.062526099;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
DR EMBL; AE009955; AAL96849.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBAPROTEIN.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
```

```
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match 51.8%; Score 1093; DB 2; Length 398;
Best Local Similarity 53.8%; Pred. No. 3.5e-41;
Matches 233; Conservative 58; Mismatches 106; Indels 36; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSATTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
Db 1 MKKRILSAVLVSGVTLSATTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQSNATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQSNATSY 120
QY 58 QVSSLSQSEQDKLTARNTLEALSKEPEQEI KALTSQIVARNEKLNQOARSAYKNNETSGY 117
Db 58 QVSSLSQSEQDKLTARNTLEALSKEPEQEI KALTSQIVARNEKLNQOARSAYKNNETSGY 117
QY 121 INAINSKSVSDAINRVSAIREVWSANEKMLQOEODKAAVEQKQOENQAAINTVAANQE 180
Db 121 INAINSKSVSDAINRVSAIREVWSANEKMLQOEODKAAVEQKQOENQAAINTVAANQE 180
QY 118 INALLNSKISIDVNRVLRVAINRAVSANAKLLEQKADKVSLEEKQAANTAIANTAANMA 177
Db 118 INALLNSKISIDVNRVLRVAINRAVSANAKLLEQKADKVSLEEKQAANTAIANTAANMA 177
QY 181 TTAQNTNALNTQQAQLEAAQLNLQALBELTTAQQOKATLVQAQKAAAEPAARCAAAQAABEA 240
Db 181 TTAQNTNALNTQQAQLEAAQLNLQALBELTTAQQOKATLVQAQKAAAEPAARCAAAQAABEA 240
QY 178 MAEENQNTLRTQANLEAATANLALQLASATEDKANLVAQKEAAEKAAAEALAEQAQAKV 237
Db 178 MAEENQNTLRTQANLEAATANLALQLASATEDKANLVAQKEAAEKAAAEALAEQAQAKV 237
QY 241 KAAAEAKALOEAQAQAQAANNNNTQATDASDQAAAAADNTQAAOTGDDTSDQSAQAQVNN 300
Db 241 KAAAEAKALOEAQAQAQAANNNNTQATDASDQAAAAADNTQAAOTGDDTSDQSAQAQVNN 300
QY 238 KA-----QEQA-----QQAASVEAKSAITPAQATPAQAQSSNA 272
Db 238 KA-----QEQA-----QQAASVEAKSAITPAQATPAQAQSSNA 272
QY 301 SDOESTTATAAOPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNVWNGGQWA 360
Db 301 SDOESTTATAAOPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNVWNGGQWA 360
QY 273 IEPAALTAPAA-PSARPQT-----SYDSSNTYPVQCTWGAKSLAPWAGNNGGQWA 325
Db 273 IEPAALTAPAA-PSARPQT-----SYDSSNTYPVQCTWGAKSLAPWAGNNGGQWA 325
QY 361 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTVGVQ-GGQIQVOEANYAGNQSIGNYRG 419
Db 361 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTVGVQ-GGQIQVOEANYAGNQSIGNYRG 419
QY 326 YSAQAAGYRTGTPMVGAIVVNDGGYGHVAVVVEVQSASSIRVMESSYSGRYIADHRG 385
Db 326 YSAQAAGYRTGTPMVGAIVVNDGGYGHVAVVVEVQSASSIRVMESSYSGRYIADHRG 385
QY 420 WFNPGSVSVIYPN 432
Db 420 WFNPGSVSVIYPN 432
QY 386 WFNPTGVTFIYPH 398
Db 386 WFNPTGVTFIYPH 398

RESULT 12
QBP318_STRP3 PRELIMINARY; PRT; 398 AA.
AC QBP318; Q7CFL7;
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 13-SEP-2005 (TremBLrel. 31, Last annotation update)
DE Putative secreted protein.
GN OrderedLocusNames=SP60015, SpyM3_0014;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301448;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SSI-1 / Serotype M3;
RX MEDLINE=22133808; PubMed=12129345; DOI=10.1101/gr.1096703;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution.";
RL Genome Res. 13:1042-1055 (2003).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=152298499;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus;
RT Bacterium Streptococcus thermophilus.";
```

```
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083 (2002).
DR EMBL; BA000034; BAC63110.1; -; Genomic DNA.
DR EMBL; AB014136; AM78621.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match 51.8%; Score 1093; DB 2; Length 398;
Best Local Similarity 53.8%; Pred. No. 3.5e-41;
Matches 233; Conservative 58; Mismatches 106; Indels 36; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSATTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
Db 1 MKKRILSAVLVSGVTLSATTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQSNATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQSNATSY 120
QY 58 QVSSLSQSEQDKLTARNTLEALSKEPEQEI KALTSQIVARNEKLNQOARSAYKNNETSGY 117
Db 58 QVSSLSQSEQDKLTARNTLEALSKEPEQEI KALTSQIVARNEKLNQOARSAYKNNETSGY 117
QY 121 INAINSKSVSDAINRVSAIREVWSANEKMLQOEODKAAVEQKQOENQAAINTVAANQE 180
Db 121 INAINSKSVSDAINRVSAIREVWSANEKMLQOEODKAAVEQKQOENQAAINTVAANQE 180
QY 118 INALLNSKISIDVNRVLRVAINRAVSANAKLLEQKADKVSLEEKQAANTAIANTAANMA 177
Db 118 INALLNSKISIDVNRVLRVAINRAVSANAKLLEQKADKVSLEEKQAANTAIANTAANMA 177
QY 181 TTAQNTNALNTQQAQLEAAQLNLQALBELTTAQQOKATLVQAQKAAAEPAARCAAAQAABEA 240
Db 181 TTAQNTNALNTQQAQLEAAQLNLQALBELTTAQQOKATLVQAQKAAAEPAARCAAAQAABEA 240
QY 178 MAEENQNTLRTQANLEAATANLALQLASATEDKANLVAQKEAAEKAAAEALAEQAQAKV 237
Db 178 MAEENQNTLRTQANLEAATANLALQLASATEDKANLVAQKEAAEKAAAEALAEQAQAKV 237
QY 241 KAAAEAKALOEAQAQAQAANNNNTQATDASDQAAAAADNTQAAOTGDDTSDQSAQAQVNN 300
Db 241 KAAAEAKALOEAQAQAQAANNNNTQATDASDQAAAAADNTQAAOTGDDTSDQSAQAQVNN 300
QY 238 KA-----QEQA-----QQAASVEAKSAITPAQATPAQAQSSNA 272
Db 238 KA-----QEQA-----QQAASVEAKSAITPAQATPAQAQSSNA 272
QY 301 SDOESTTATAAOPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNVWNGGQWA 360
Db 301 SDOESTTATAAOPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNVWNGGQWA 360
QY 273 IEPAALTAPAA-PSARPQT-----SYDSSNTYPVQCTWGAKSLAPWAGNNGGQWA 325
Db 273 IEPAALTAPAA-PSARPQT-----SYDSSNTYPVQCTWGAKSLAPWAGNNGGQWA 325
QY 361 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTVGVQ-GGQIQVOEANYAGNQSIGNYRG 419
Db 361 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTVGVQ-GGQIQVOEANYAGNQSIGNYRG 419
QY 326 YSAQAAGYRTGTPMVGAIVVNDGGYGHVAVVVEVQSASSIRVMESSYSGRYIADHRG 385
Db 326 YSAQAAGYRTGTPMVGAIVVNDGGYGHVAVVVEVQSASSIRVMESSYSGRYIADHRG 385
QY 420 WFNPGSVSVIYPN 432
Db 420 WFNPGSVSVIYPN 432
QY 386 WFNPTGVTFIYPH 398
Db 386 WFNPTGVTFIYPH 398

RESULT 13
QSM6K4_STRT2 PRELIMINARY; PRT; 474 AA.
AC QSM6K4;
DT 01-FEB-2005 (TremBLrel. 29, Created)
DT 01-FEB-2005 (TremBLrel. 29, Last sequence update)
DT 13-SEP-2005 (TremBLrel. 31, Last annotation update)
DE Glucan binding protein (PcSB).
GN Name=PcSB; OrderedLocusNames=stu0022;
OS Streptococcus thermophilus (strain ATCC BAA-250 / LMG 18311).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=264199;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15543133; DOI=10.1038/nbt1034;
RA Bolotin A., Quinquis B., Renault P., Sorokin A., Ehrlich S.D.,
RA Kulakauskas S., Lapidus A., Goltzman E., Mazur M., Pusch G.D.,
RA Fongstein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D.,
RA Ngui K., Masuy D., Hancy F., Burtreau S., Boutry M., Delcours J.,
RA Goffeau A., Hols P.;
RT "Complete sequence and comparative genome analysis of the dairy
RT bacterium Streptococcus thermophilus.";
```

```

RL Nat. Biotechnol. 22:1554-1558(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=LMG 18311;
RA Borges F., Layec S., Thibessard A., Fernandez A., Gintz B., Hols P.,
RA Decaris B., Leblond-Bourget N.;
RT "cse, a Chimeric and Variable Gene, Encodes an Extracellular Protein
RT Involved in Cellular Segregation in Streptococcus thermophilus.";
RL J. Bacteriol. 187:2737-2746(2005).
DR EMBL; CP000023; AAV59752.1; -; Genomic DNA.
DR EMBL; AY730643; AAW82375.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 474 AA; 48142 MW; 1BAC6A9E0A0A200D CRC64;

Query Match 51.8%; Score 1092.5; DB 2; Length 474;
Best Local Similarity 51.7%; Pred. No. 4.4e-41;
Matches 247; Conservative 73; Mismatches 107; Indels 51; Gaps 12;

QY 1 MKKRILSAVLVSGVTLSATLSAISKADDDFAQIASQDSKINNLTAAQQAQAQVNTTQ 60
DB 1 MKKRILSAVLVSGVTLSA---ASVHAEDYDSQIAATNNAISNLASQQAQAQVATIQS 57

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQQAARSQAQNAATSY 120
DB 58 QVSTLRTQKTELEAKNAELEKVSADLESEIQELSSKIVARQDSLAKQAARSQAQNTATSY 117

QY 121 INAINSKSVSDAINRVSAIREVVSANEKMLQQQODKAAVEKQOENQAQNTVAANO 180
DB 118 INSILNSKISSEAITRITAISKVVTANNLLTKQESDQKELAAKQOENQAQNTIAANKS 177

QY 181 TTAQNTNALNTQOALEAQLNLOAELTTAQDOKATLVAQKAAAEAAARQAAA---AQA 236
DB 178 ELETTEAGLTQQAELEAAQVTLAEALATAQNEKTSLSVSAKSTAESVAASTAASVAQSOA 237

QY 237 AAEAKA-----AAEAKALQEAQAQAQAANNNTQATDASDQQAQAADNTQAAQTG 287
DB 238 IAESSEATAQVVASSEATSVASSEVAATSEAVQSETPVSEIS-TASEAAQEPASSETS 296

QY 288 DSTDQSAQAQVNN-----SDQESTTATA-----AQSASSA----- 318
DB 297 EVQPESAAPVSEAPASVAPVATSEAPATSEAPASVAPVATSEAPVSEAPVSAAPV 356

QY 328 -----SSANTYPAGCTWGVKSLAPWGVNGGQWAAASAAAAGYRVGTSAGAVAVW 382
DB 357 KVSAASTPNTYPVGOCTWGVKSLAPWAGNNGWNAKNWLTASQAAGHSVGTTPVAGAI 416

QY 383 -ND-GGYGHVAVYTVGQ-GQIQVEANYAGNOSIGNYRGWFPN-----GSVSYIYP 431
DB 417 PNDGGYGHVAVYTVSAGSANSIQVMESYAGNMSISNYRGTFDPTSSAHGGSVFIYP 474

RESULT 14
QSM212_STR11 PRELIMINARY; PRT; 485 AA.
AC QSM212;
DT 01-FEB-2005 (TremBLrel. 29, Created)
DT 01-FEB-2005 (TremBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TremBLrel. 29, Last annotation update)
DE Glucan binding protein.
GN Name=pcsb; OrderedLocusNames=st0022;
OS Streptococcus thermophilus (strain CNR3 1066).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=299768;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed15543133; DOI=10.1038/nbt1034;
RA Bolotin A., Quinquis B., Renault P., Sorokin A., Ehrlich S.D.,

```

```

RA Kulakauskas S., Lapidus A., Goltzman E., Mazur M., Pusch G.D.,
RA Fonstein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D.,
RA Ngui K., Masuy D., Hancy F., Burteau S., Boutry M., Delcours J.,
RA Goffeau A., Hols P.;
RT "Complete sequence and comparative genome analysis of the dairy
RT bacterium Streptococcus thermophilus.";
RL Nat. Biotechnol. 22:1554-1558(2004).
DR EMBL; CP000024; AAV61641.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 485 AA; 49165 MW; 8665E4B1FC6CF7CA CRC64;

Query Match 51.8%; Score 1092; DB 2; Length 485;
Best Local Similarity 50.9%; Pred. No. 4.8e-41;
Matches 249; Conservative 71; Mismatches 107; Indels 62; Gaps 12;

QY 1 MKKRILSAVLVSGVTLSATLSAISKADDDFAQIASQDSKINNLTAAQQAQAQVNTTQ 60
DB 1 MKKRILSAVLVSGVTLSA---ASVHAEDYDSQIAATNNAISNLASQQAQAQVATIQS 57

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQQAARSQAQNAATSY 120
DB 58 QVSTLRTQKTELEAKNAELEKVSADLESEIQELSSKIVARQDSLAKQAARSQAQNTATSY 117

QY 121 INAINSKSVSDAINRVSAIREVVSANEKMLQQQODKAAVEKQOENQAQNTVAANO 180
DB 118 INSILNSKISSEAITRITAISKVVTANNLLTKQESDQKELAAKQOENQAQNTIAANKS 177

QY 181 TTAQNTNALNTQOALEAQLNLOAELTTAQDOKATLVAQKAAAEAAARQAAA---AQA 236
DB 178 ELETTEAGLTQQAELEAAQVTLAEALATAQNEKTSLSVSAKSTAESVAASTAASVAQSOA 237

QY 237 AAEAKA-----AAEAKALQEAQAQAQAANNNTQATDASDQQAQAADNTQAAQTG 287
DB 238 IAESSEATAQVVASSEATSVASSEVAATSEAVQSETPVSEIS-TASEAAQEPASSETS 296

QY 288 DSTDQSAQAQVNN-----SDQESTTATA-----AQSASSA----- 318
DB 297 EVQPESAAPVSEAPASVAPVATSEAPATSEAPASVAPVATSEAPVSEAPVSAAPV 356

QY 319 -STAANAANT-----SSANTYPAGCTWGVKSLAPWGVNGGQWAAASAAAAGYRVG 371
DB 357 TSEAPAAETHKVSAASTPNTYPVGOCTWGVKSLAPWAGNNGWNAKNWLTASQAAGHSV 416

QY 372 STPSAGAVAVW-ND-GGYGHVAVYTVGQ-GQIQVEANYAGNOSIGNYRGWFPN----- 423
DB 417 TTPVAGAIWVNDGGYGHVAVYTVSAGSANSIQVMESYAGNMSISNYRGTFDPTSSAH 476

QY 424 -GSVSYIYP 431
DB 477 GGSVFIYP 485

RESULT 15
Q9A1Z8_STRPY PRELIMINARY; PRT; 398 AA.
AC Q9A1Z8; Q7BH59;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 10-MAY-2005 (TremBLrel. 30, Last annotation update)
DE Secreted protein Siba precursor.
GN OrderedLocusNames=SPY0019;
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1314;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;

```



```

RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Fagan P.K., Reinscheid D., Gottschalk B., Chhatwal G.S.;
RT "Identification and characterization of a novel secreted protein from
RT group A streptococcus.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB06474; AAK33158.1; -; Genomic DNA.
DR EMBL; AF319999; AAL73135.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome; signal.
FT SIGNAL 1
FT SIGNAL 23
SQ SEQUENCE 398 AA; 41899 MW; 28A9B3F7195E969B CRC64;

Query Match 51.5%; Score 1087; DB 2; Length 398;
Best Local Similarity 53.6%; Pred. No. 6.6e-41;
Matches 232; Conservative 58; Mismatches 107; Indels 36; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSSNTLSAISKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLGAATT---VGAEDLSTKIAKQDSIIISNLTTEKAAQNVSAQQA 57
QY 61 QVSALQTQQAELQAEORLEAQSATLGGQIOITLSSKIVARNESLKQQAASAKSNAATSY 120
DB 58 QVSSLSQSQDKLTARTLEALSKEFEQEIKALTSQIVARNEKLNQAESAYKNETSGY 117
QY 121 INAIINSKSVDAINRVSAIREVVSANERKMLQQQODKAAVEQKQENQAQAAINTVAANQE 180
DB 118 INALLNSKISIDVNRKLVAINRAVSANAKLLRQKADKVSLEKQAANQTAINTIAANWA 177
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAEILTAAQDQKATLVAQKAAEEAARQAAAQAAAEA 240
DB 178 MAEENQNTLRTOQANLVAATANLALQLASATEDKANLVAQKEAAEKAAAEALAEQAQKV 237
QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAOTGDSDDQAAQAVNN 300
DB 238 KA-----QEQA-----QQAASVEAAKSALTAPAQATPAQSSNA 272
QY 301 SDQESTTATAQPSASSASTAAVAANTSANTYPAGQCTWGVKSLAPWVGYNGGQOWA 360
DB 273 IEPAALTAPAA-PSAGPQT-----SYDSSNTYPPVQCCTWGAKSLAPWAGNNGGQOWA 325
QY 361 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVVTGVQ-GGQIQOVBANYAGNQSIGNYRG 419
DB 326 YSAQAAGYRTGSTPMVGAIAVNDGGYGHVAVVVEVQSASSIRVMESNYSGRQYIADHRG 385
QY 420 WFNPGSVSVIYPN 432
DB 386 WFNPTGVTFIYPH 398
```

Search completed: February 15, 2006, 18:09:15  
Job time : 127.762 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:09:37 ; Search time 30.0232 Seconds  
(without alignments)  
1189.611 Million cell updates/sec

Title: US-10-797-821-31  
Perfect score: 2110  
Sequence: 1 MKRILSAVLVSGVTLSAT.....SIGNRGWFMFGSVIYPN 432

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA:\*  
1: /cgn2\_6/prodata/1/iaa/5\_COMB.pep.\*  
2: /cgn2\_6/prodata/1/iaa/6\_COMB.pep.\*  
3: /cgn2\_6/prodata/1/iaa/H\_COMB.pep.\*  
4: /cgn2\_6/prodata/1/iaa/PCUS\_COMB.pep.\*  
5: /cgn2\_6/prodata/1/iaa/RE\_COMB.pep.\*  
6: /cgn2\_6/prodata/1/iaa/baCkfilesl.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description          |
|------------|-------|-------------|--------|-------|----------------------|
| 1          | 1059  | 50.2        | 399    | 2     | US-09-107-433-3230   |
| 2          | 1058  | 50.1        | 392    | 2     | US-09-583-110-4374   |
| 3          | 681.5 | 32.3        | 461    | 1     | US-08-186-222-2      |
| 4          | 584.5 | 27.7        | 525    | 2     | US-09-107-532A-5095  |
| 5          | 561   | 26.6        | 210    | 2     | US-09-222-938A-67    |
| 6          | 539   | 25.1        | 497    | 2     | US-09-134-000C-5990  |
| 7          | 500.5 | 23.7        | 449    | 2     | US-09-071-035-482    |
| 8          | 500.5 | 23.7        | 449    | 2     | US-10-206-576-482    |
| 9          | 500.5 | 23.7        | 450    | 2     | US-09-134-000C-5714  |
| 10         | 452.5 | 21.4        | 422    | 2     | US-09-071-035-484    |
| 11         | 452.5 | 21.4        | 422    | 2     | US-10-206-576-484    |
| 12         | 283.5 | 13.4        | 469    | 2     | US-09-489-039A-13565 |
| 13         | 237   | 11.2        | 264    | 2     | US-09-134-001C-5035  |
| 14         | 236   | 11.2        | 257    | 2     | US-09-170-279-3244   |
| 15         | 236   | 11.2        | 267    | 2     | US-09-134-001C-4539  |
| 16         | 235   | 11.1        | 477    | 2     | US-09-302-540-11649  |
| 17         | 232   | 11.0        | 2310   | 2     | US-09-874-923-120    |
| 18         | 226   | 10.7        | 1236   | 2     | US-09-769-787-109    |
| 19         | 222.5 | 10.5        | 270    | 2     | US-09-134-001C-5441  |
| 20         | 214.5 | 10.2        | 266    | 2     | US-09-134-001C-5453  |
| 21         | 214.5 | 10.2        | 468    | 2     | US-09-328-352-6321   |
| 22         | 214   | 10.1        | 149    | 2     | US-09-710-279-1682   |
| 23         | 214   | 10.1        | 157    | 2     | US-09-710-279-2870   |
| 24         | 213   | 10.1        | 440    | 2     | US-08-302-756F-35    |
| 25         | 212.5 | 10.1        | 655    | 2     | US-09-902-540-10005  |
| 26         | 212   | 10.0        | 610    | 2     | US-09-336-447A-11    |
| 27         | 212   | 10.0        | 610    | 2     | US-09-952-267B-11    |

|    |       |     |      |   |                      |                   |
|----|-------|-----|------|---|----------------------|-------------------|
| 28 | 208.5 | 9.9 | 1566 | 1 | US-08-687-956A-23    | Sequence 23, Appl |
| 29 | 206   | 9.8 | 679  | 2 | US-09-489-039A-12307 | Sequence 12307, A |
| 30 | 203.5 | 9.6 | 639  | 2 | US-09-902-540-14908  | Sequence 14908, A |
| 31 | 203.5 | 9.6 | 1070 | 2 | US-09-902-540-13861  | Sequence 13861, A |
| 32 | 202.5 | 9.6 | 955  | 1 | US-08-006-676B-1     | Sequence 1, Appl  |
| 33 | 202.5 | 9.6 | 955  | 1 | US-08-282-845-2      | Sequence 2, Appl  |
| 34 | 202.5 | 9.6 | 955  | 1 | US-08-428-414A-3     | Sequence 3, Appl  |
| 35 | 202.5 | 9.6 | 955  | 4 | PCT-US94-00324-1     | Sequence 1, Appl  |
| 36 | 197.5 | 9.4 | 1024 | 2 | US-09-270-767-44973  | Sequence 44973, A |
| 37 | 196.5 | 9.3 | 883  | 2 | US-09-489-039A-12755 | Sequence 12755, A |
| 38 | 195.5 | 9.3 | 443  | 1 | US-08-795-475-6      | Sequence 6, Appl  |
| 39 | 195.5 | 9.3 | 443  | 2 | US-08-325-278B-6     | Sequence 6, Appl  |
| 40 | 193   | 9.1 | 2107 | 2 | US-09-949-016-7646   | Sequence 7646, Ap |
| 41 | 193   | 9.1 | 2107 | 2 | US-09-949-016-7647   | Sequence 7647, Ap |
| 42 | 193   | 9.1 | 2115 | 2 | US-09-296-662-33     | Sequence 33, Appl |
| 43 | 191.5 | 9.1 | 2101 | 1 | US-08-466-390-4      | Sequence 4, Appl  |
| 44 | 191.5 | 9.1 | 2101 | 1 | US-08-470-950-4      | Sequence 4, Appl  |
| 45 | 191.5 | 9.1 | 2101 | 1 | US-08-467-781-4      | Sequence 4, Appl  |

ALIGNMENTS

RESULT 1

US-09-107-433-3230

; Sequence 3230, Application US/09107433  
; Patent No. 6800744

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE

; THERAPEUTICS

; NUMBER OF SEQUENCES: 5206

; CORRESPONDENCE ADDRESS:

; ADDRESS: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; City: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-ROM ISO9660

; COMPUTER: <Unknown>

; OPERATING SYSTEM: <Unknown>

; SOFTWARE: <Unknown>

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,433

; FILING DATE: 30-Jun-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/ 085131

; FILING DATE: May 12, 1998

; APPLICATION NUMBER: 60/051553

; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Arinello, Pamela Deneke

; REGISTRATION NUMBER: 40,489

; REFERENCE/DOCKET NUMBER: GTC-011

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781)893-5007

; TELEFAX: (781)893-8277

; INFORMATION FOR SEQ ID NO: 3230:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 399 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: YES

; ORIGINAL SOURCE:

; ORGANISM: Streptococcus pneumoniae

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (8) LOCATION 1...399

; SEQUENCE DESCRIPTION: SEQ ID NO: 3230:

US-09-107-433-3230

Query Match 50.2%; Score 1059; DB 2; Length 399;  
Best Local Similarity 50.0%; Pred. No. 3.6e-73;  
Matches 218; Conservative 75; Mismatches 91; Indels 52; Gaps 4;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAISKADDFDAQLASODSKINNLTAAQQAQAVNTIQ 60  
DB 8 MKKKILASLLSTVWMSQVAVTTAAETTTDKIAAQDNKISNLTAAQQAQAVNTIQ 67

QY 61 QVSALQTOQAEQLAENRLEAQSATLIGQOITLSSKIVARNESLKKQARSQAQNAATSY 120  
DB 68 QVSAIQAESNLQAEENDRLQAESKKLEGEITELSKNIVSRNOSLEKQARSQAQNGAVTSY 127

QY 121 INAINSKSVDAINRVSAIREVVSANERKMLQQOQODKAAVEKQKQENQAAINTVAANO 180  
DB 128 INTIVNSKSIETASIRVAAMSEIVSANNKMLEEQKADKKAISEKQVANNDAINTVIAN 187

QY 181 TTAQNTNALNTQOQAELEAQLNLQAEELTTAODOKATLVAQKAAAEAEARQAAAAA 240  
DB 188 KIADDAQALTTTQAELEKAAELSLAAEKATAGEKASLLEQKAAAEAEARAAVAEAY 247

QY 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTQSAQA 300  
DB 248 KRASQQQSV-----LASANTNLTAQVQAVSESAAAAPVRKVRPT----- 286

QY 301 SDOESTTATAAOPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWVGNWGGQWA 360  
DB 287 -----YSTNASSYPIGECTGWGKTLAPWAGDYWGNGAQWA 321

QY 361 ASAAAAGYRVGTPSAGAVAVMNDGGYGHVAVVTGQV-GQIQVOEANYAGNOSIGNY 419  
DB 322 TSSAAAAGFRTGSTPOVGAIAACWNDGGYGHVAVVTAVESTTRIQVSESYNAGNRTIGNHRG 381

QY 420 WFNPP-----GSVSYIY 430  
DB 382 WFNPTTSEGFTVIY 397

RESULT 2  
US-09-583-110-4374  
; Sequence 4374, Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE REFERENCE: Nucleonidae for Diagnostics and Therapeutics  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; CURRENT FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 4374  
; LENGTH: 392  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae

US-09-583-110-4374  
Query Match 50.1%; Score 1058; DB 2; Length 392;  
Best Local Similarity 50.0%; Pred. No. 4.2e-73;  
Matches 218; Conservative 74; Mismatches 92; Indels 52; Gaps 4;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAISKADDFDAQLASODSKINNLTAAQQAQAVNTIQ 60  
DB 1 MKKKILASLLSTVWMSQVAVTTAAETTTDKIAAQDNKISNLTAAQQAQAVNTIQ 60

QY 61 QVSALQTOQAEQLAENRLEAQSATLIGQOITLSSKIVARNESLKKQARSQAQNAATSY 120  
DB 68 QVSAIQAESNLQAEENDRLQAESKKLEGEITELSKNIVSRNOSLEKQARSQAQNGAVTSY 127

DB 61 QVSAIQAESNLQAEENDRLQAESKKLEGEITELSKNIVSRNOSLEKQARSQAQNGAVTSY 120  
QY 121 INAINSKSVDAINRVSAIREVVSANERKMLQQOQODKAAVEKQKQENQAAINTVAANO 180  
DB 128 INTIVNSKSIETASIRVAAMSEIVSANNKMLEEQKADKKAISEKQVANNDAINTVIAN 180  
QY 181 TTAQNTNALNTQOQAELEAQLNLQAEELTTAODOKATLVAQKAAAEAEARQAAAAA 240  
DB 188 KIADDAQALTTTQAELEKAAELSLAAEKATAGEKASLLEQKAAAEAEARAAVAEAY 240  
QY 241 KAAAEAKALQEQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGSDTQSAQA 300  
DB 248 KRASQQQSV-----LASANTNLTAQVQAVSESAAAAPVRKVRPT----- 279  
QY 301 SDOESTTATAAOPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWVGNWGGQWA 360  
DB 287 -----YSTNASSYPIGECTGWGKTLAPWAGDYWGNGAQWA 314  
QY 361 ASAAAAGYRVGTPSAGAVAVMNDGGYGHVAVVTGQV-GQIQVOEANYAGNOSIGNY 419  
DB 322 TSSAAAAGFRTGSTPOVGAIAACWNDGGYGHVAVVTAVESTTRIQVSESYNAGNRTIGNHRG 374  
QY 420 WFNPP-----GSVSYIY 430  
DB 375 WFNPTTSEGFTVIY 390

RESULT 3  
US-08-186-222-2  
; Sequence 2, Application US/08186222  
; Patent No. 5559007  
; GENERAL INFORMATION:  
; APPLICANT: Suri, Bruno  
; APPLICANT: Schmitz, Albert  
; TITLE OF INVENTION: Bacterial Vectors  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/186,222  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,205  
FILING DATE: 19-MAR-1991  
APPLICATION NUMBER: GB 9006400.7  
FILING DATE: 22-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Villamizar, JoAnn  
REGISTRATION NUMBER: 30,598  
REFERENCE/DOCKET NUMBER: 4-17994/A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (914)785-7121  
TELEFAX: (914)347-5769  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 461 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-186-222-2  
Query Match 32.3%; Score 681.5; DB 1; Length 461;

```
Best Local Similarity 35.3%; Pred. No. 3.4e-44;
Matches 170; Conservative 106; Mismatches 136; Indels 69; Gaps 14;

QY 1 MKKRILSAVLVSGVTLSATLSAIDKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60
Db 1 MKKKIISAILMTSVLSAAPLSGVYADT-NSDIAQDATISSAQSAKAAQAQVDSLOS 59
QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQIOTLSSKIVARNESLKKQARSQAQSNAAATSY 120
Db 60 KVDLSLQOKQTSTKAQIAKTESERKALNAQIATLINESIKERTKTLEAQAARSQAQVNSATNY 119
QY 121 INAINSKSVSDAINRVSAIREVVSANEMKLOOEQDKAAVEKQKQENQAAINTVAANQE 180
Db 120 MDVAVNSKSLTDVIOKVTAIATVSSANKQMLEQKEQKELSKSETVKKYNNQFVLSLQ 179
QY 181 TTAQNTNALNTQQAOLEEAQLNLQAEELTTAQQOKATLVAKAAAEAAEAQAQAAAAA 240
Db 180 SLDSQAQELTSQAQELKVATLNYQATIAQDKQALLDEKAAAEKAAQEAQAKQAAYEA 239
QY 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQQAQAANNNTQAAQGTGSDTQ-----SAA 295
Db 240 Q-----QKEAAQAQAATAATAKAVEAA-TSSASASSQAQPVSTSTDTNTTSNASAS 290
QY 296 QAVNSDQESTTATAAQPSSASSASTAAVANTSS-----ANTY 333
Db 291 NSSNSSNSSSSSSSSSSSSSSSSSSSSSSNAGGNTNSTGTCGTGTTGGSGINSPIGNPY 350
QY 334 PAQOCT--WG-----VKSLAPVGVNGYGQWAAASAAAG--YRVGSTPSAGAVA 380
Db 351 AVGGCTDYQWYPAAGIYIRIMP-----GNGGQWASNGPAQGVHLVVGAAP--GVIA 402
QY 381 V-----WNDGGYGHVAVTGVQV-GQIQVQEAANYAGNQSIGNVYRGWFGSVSYIYP 431
Db 403 SSFSADPVGYSNPGYHVAIVKSVNSDGTITIKGGY-GTTWGHGR-TVSASGVTFLMP 460
QY 432 N 432
Db 461 N 461

RESULT 4
US-09-107-532A-5095
; Sequence 5095, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSER: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
```

```
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5095:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 525 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (B) LOCATION 1...525
; SEQUENCE DESCRIPTION: SEQ ID NO: 5095:
US-09-107-532A-5095

Query Match 27.7%; Score 584.5; DB 2; Length 525;
Best Local Similarity 30.4%; Pred. No. 1.1e-36;
Matches 160; Conservative 92; Mismatches 163; Indels 111; Gaps 14;

QY 1 MKKRILSAVLVSGVTLSATLSAIDKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60
Db 2 VKKSIISAVVWCSMTLTAVASPIRAAADFDQIQODQKIADLKNQKQADASQIDALES 61
QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQIOTLSSKIVARNESLKKQARSQAQSNAAATSY 120
Db 62 QVSEINTQAQDLAKQDTLRQESAQLVKDIADLQERIEKREDTIQKQAREAQVNTSSNY 121
QY 121 INAINSKSVSDAINRVSAIREVVSANEMKLOOEQDKAAVEKQKQENQAAINTVAANQE 180
Db 122 IDAVLNADSLADAIGRVQAMTTMVKANNDLMEQKQDKKAVEDKKAENDAKLKEAENQA 181
QY 181 TTAQNTNALNTQQAOLEEAQLNLQAEELTTAQQOKATLVAKA-----AAEE 226
Db 182 ALESQKGLLSKQADLNLVKTSLAAEQATAEKDKADLNFQKAEAEQARIREQQRLAEQ 241
QY 227 AARQAAAAQAQAAEAKAAEAKALQEAQAQAQAQAANNNTQATDAS-----DQAAAAAD 279
Db 242 ARQAAQAEKAEKEAREQAEAEA--QATQASSTAQSSATEESSATQSSMTTESSSATQSS 298
QY 280 NTQAAQGTGSDTQSAQAQAVNNSDQESTTA-----TAAQPSASSASTAAVAANTSS 329
Db 299 ATEESTTPESSTESSTAPESSESTTAPESSESTTTPESSESTTTPESSESTTTPESSTTTE 358
QY 330 ANTYPA-----GQCTWGVKSLAP----- 347
Db 359 ESTTPATPTPSTDQSDVDGTNGT-GSSTPATPTPTTPEQPKVTPAPAPSGSVNGAIVAE 417
QY 348 ---WVGN---YWGNG-----GOWAASAAAAAGYRVG-STPSAGA 378
Db 418 AVKYIGTPYVWGGKDPSCGFDSCGFYVYVMTGTGRDIGMTVPQESAGTKISVQAKAGD 477
QY 379 VAVW-NDGGYGHVAVTGVQGGQIQVQEAANYAGNQSIGNVYRGWFGVFP 423
Db 478 LLFWGSGQGTTHVAIALG--GQQ-YIHAPQPGESVKVGSVQ-WFAP 519

RESULT 5
US-09-222-938A-67
; Sequence 67, Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222,938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 67
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-222-938A-67

Query Match          26.6%; Score 561; DB 2; Length 210;
Best Local Similarity 55.2%; Pred. No. 2e-35;
Matches 116; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSATLTSALIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
DB 1 MKKRILASLLSTVMVSVQAVLTTHAETDDKIAADNKISNLTAQQAQAQVNDIQE 60

QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQIOTLSSKIVARNESLKOQARSQAQSNAAATSY 120
DB 61 QVSALQAEQSNLQAEANDRLQAESKKLEGEITELSKNIVSRNQSLKQARSQAQTNQAVTSY 120

QY 121 INAIINSKSVDAINRVSAREVVSANERKMLQOQEQDQKAAVEQKQEQENQAANTVAANOE 180
DB 121 INTIINSKSITPAISRAVAMSIVSANNKMLQEQKADKKAISEKQVANNDAINTVIANQQ 180

QY 181 TTAQNTNALNTQQAOLEAQLNLQAEELTTA 210
DB 181 KLADDAQAALTTKQAEKAAELSLAAEKATS 210

RESULT 6
US-09-134-000C-5990
; Sequence 5990, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134, 000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055, 778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5990
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5990

Query Match          25.1%; Score 529; DB 2; Length 497;
Best Local Similarity 31.3%; Pred. No. 1.8e-32;
Matches 152; Conservative 83; Mismatches 174; Indels 76; Gaps 12;

QY 1 MKKRILSAVLVSGVTLSSATLTSALIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
DB 16 LKSVLSALMVCISILTSVALPSAFADEYDFYKIQQDQKINALTSQMSDAEAKVAIEN 75

QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQIOTLSSKIVARNESLKOQARSQAQSNAAATSY 120
DB 76 DMVETAKQIDTTLTAKKNLSSESVKLYSEISDLNVIQKREVQMTKQARDVQVNGQSDSI 135

QY 121 INAIINSKSVDAINRVSAREVVSANERKMLQOQEQDQKAAVEQKQEQENQAANTVAANOE 180
DB 136 IDAVIDADSVAIDAGRQVAVSTMNANLLEQQEQDKATVEKTKRNVEKQIAELEAATK 195

QY 181 TTAQNTNALNTQQAOLEAQLNLQAEELTTA 210
DB 196 ELNDKTESLTKLKIQEVAKNDLEAQRSEEQKDGFIKQKXEAELAEQARQRAAK 255

QY 235 -----QAAAEAKAAAEAKALQEQAAQAAQAAANNNTQATDASQQAADNTQAAQTGDS 290
DB 256 KAEQQAQAQAQAQ- KAAAEQAKATKAA-----NEAASAAAEKAA-----TPVVSSTTT 306

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 67
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-222-938A-67

Query Match          26.6%; Score 561; DB 2; Length 210;
Best Local Similarity 55.2%; Pred. No. 2e-35;
Matches 116; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSATLTSALIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
DB 1 MKKRILASLLSTVMVSVQAVLTTHAETDDKIAADNKISNLTAQQAQAQVNDIQE 60

QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQIOTLSSKIVARNESLKOQARSQAQSNAAATSY 120
DB 61 QVSALQAEQSNLQAEANDRLQAESKKLEGEITELSKNIVSRNQSLKQARSQAQTNQAVTSY 120

QY 121 INAIINSKSVDAINRVSAREVVSANERKMLQOQEQDQKAAVEQKQEQENQAANTVAANOE 180
DB 121 INTIINSKSITPAISRAVAMSIVSANNKMLQEQKADKKAISEKQVANNDAINTVIANQQ 180

QY 181 TTAQNTNALNTQQAOLEAQLNLQAEELTTA 210
DB 181 KLADDAQAALTTKQAEKAAELSLAAEKATS 210

RESULT 7
US-09-071-035-482
; Sequence 482, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brooks
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 482:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; LENGTH: 449 amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-482

Query Match          23.7%; Score 500.5; DB 2; Length 449;
Best Local Similarity 30.7%; Pred. No. 2.4e-30;
Matches 139; Conservative 88; Mismatches 175; Indels 51; Gaps 12;

QY 1 MKKRILSAVLVSGVTLSSATLTSALIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
DB 1 VKKRILFASVLLCSLTLSAITSALADNVDKIEKNQEISSLKAKQDGLASQVSSLEA 60

QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQIOTLSSKIVARNESLKOQARSQAQSNAAATSY 120
DB 61 EVSSVFDSESMALREKQKTLKAKSEQLQEQEITNLNORIEKRNKAIKNQARDVQVNGQSTTM 120

QY 121 INAIINSKSVDAINRVSAREVVSANERKMLQOQEQDQKAAVEQKQEQENQAANTVAANOE 180
```

Db 121 LDAVLADSVDAISRQAVSTIVSANNDLMOQKEDKQAVVDKKAENEKKVKQLEATEA 180  
QY 181 TTAQNTNALNTQQAQLAEALQNLQAEELTTAQQOKATLVAQKAAAE-----E 226  
Db 181 ELETQRDQLLSQSELNVKASLALQESAESKAGLEKQKAAAEQARLAAEQKAAAE 240  
QY 227 AARQAAAAQAAAEAKAAAEKALQEQAAQAAAAAANNNTQATDASDQAAA-----AADNT 281  
Db 241 KAKQAAAAKPAKAEVK--AEAPVASSSTTEAQAPASSSS--ATESSTQQTETTTPSTDNS 296  
QY 282 QAAQTGDSSTDQSAQAQAV-----NNSDOESTTATAAQPSASSASTAAVAANTSSANTYPA 335  
Db 297 ATENTGSSSESEQVQPTTPSDNGNGGQTGGTGTPTTPTTAPSADPTINALNLVLRQSL 356  
QY 336 GOCTGWKSLAPWVGNMGNGGOWAASAAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYTG 395  
Db 357 G-----LRPVV---WDAGLAASATARAQAQVEAGGIENDH-----WSRG--DEVIAIMW 399  
QY 396 VOGGOIQV---QEAANYAGNQSIGNYRGW--FNP 424  
Db 400 AFGNSVIMAWYNETNMVTASGSG-HRDWEINFG 431

RESULT 8

US-10-206-576-482  
; Sequence 482, Application US/10206576  
; Patent No. 6913907  
; GENERAL INFORMATION:  
; APPLICANT: Choi et al.  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 497  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-R  
; COMPUTER: Dell Latitude  
; OPERATING SYSTEM: Windows 98  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/206,576  
; FILING DATE: 29-Jul-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/071,035  
; FILING DATE: 1998-05-04  
; APPLICATION NUMBER: US 60/046,655  
; FILING DATE: 1997-05-16  
; APPLICATION NUMBER: US 60/044,031  
; FILING DATE: 1997-05-06  
; APPLICATION NUMBER: US 60/066,009  
; FILING DATE: 1997-11-14  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hyman, Mark J.  
; REGISTRATION NUMBER: 46,789  
; REFERENCE/DOCKET NUMBER: PB369PD1  
; INFORMATION FOR SEQ ID NO: 482:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 449 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 482:  
US-10-206-576-482  
Query Match 23.7%; Score 500.5; DB 2; Length 449;  
Best Local Similarity 30.7%; Pred. No. 2.4e-30;  
Matches 139; Conservative 88; Mismatches 175; Indels 51; Gaps 12;

Matches 139; Conservative 88; Mismatches 175; Indels 51; Gaps 12;  
QY 1 MKKRILSAVLVSGVTLSATTLSSATLSAISKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60  
Db 1 VKKRLFASVLLCSLTLSAIAATPSIALADNVDKKIEKQOEISSLKAKQGDLASQVSSLEA 60  
QY 61 QVSALQTOQAEIQAENORLEAQSATLGGQIQITLSSKIVARNESLKQOARSQAQSNAAATSY 120  
Db 61 EVSSVFDESMAUREQKTLKAKSEQLQEQITLNLQRIEKNRAIKNQARDVQVNGOSTTM 120  
QY 121 INAIINSKSVDAINRVSAIREVSNKMLQOQEQDQKAAVEQKQOENQAAINTVAANOE 180  
Db 121 LDAVLADSVDAISRQAVSTIVSANNDLMOQKEDKQAVVDKKAENEKKVKQLEATEA 180  
QY 181 TTAQNTNALNTQQAQLAEALQNLQAEELTTAQQOKATLVAQKAAAE-----E 226  
Db 181 ELETQRDQLLSQSELNVKASLALQESAESKAGLEKQKAAAEQARLAAEQKAAAE 240  
QY 227 AARQAAAAQAAAEAKAAAEKALQEQAAQAAAAAANNNTQATDASDQAAA-----AADNT 281  
Db 241 KAKQAAAAKPAKAEVK--AEAPVASSSTTEAQAPASSSS--ATESSTQQTETTTPSTDNS 296  
QY 282 QAAQTGDSSTDQSAQAQAV-----NNSDOESTTATAAQPSASSASTAAVAANTSSANTYPA 335  
Db 297 ATENTGSSSESEQVQPTTPSDNGNGGQTGGTGTPTTPTTAPSADPTINALNLVLRQSL 356  
QY 336 GOCTGWKSLAPWVGNMGNGGOWAASAAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYTG 395  
Db 357 G-----LRPVV---WDAGLAASATARAQAQVEAGGIENDH-----WSRG--DEVIAIMW 399  
QY 396 VOGGOIQV---QEAANYAGNQSIGNYRGW--FNP 424  
Db 400 AFGNSVIMAWYNETNMVTASGSG-HRDWEINFG 431  
RESULT 9  
US-09-134-000C-5714  
; Sequence 5714, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5714  
; LENGTH: 450  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-5714  
Query Match 23.7%; Score 500.5; DB 2; Length 450;  
Best Local Similarity 30.7%; Pred. No. 2.4e-30;  
Matches 139; Conservative 88; Mismatches 175; Indels 51; Gaps 12;  
QY 1 MKKRILSAVLVSGVTLSATTLSSATLSAISKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60  
Db 2 VKKRLFASVLLCSLTLSAIAATPSIALADNVDKKIEKQOEISSLKAKQGDLASQVSSLEA 61  
QY 61 QVSALQTOQAEIQAENORLEAQSATLGGQIQITLSSKIVARNESLKQOARSQAQSNAAATSY 120  
Db 62 EVSSVFDESMAUREQKTLKAKSEQLQEQITLNLQRIEKNRAIKNQARDVQVNGOSTTM 121  
QY 121 INAIINSKSVDAINRVSAIREVSNKMLQOQEQDQKAAVEQKQOENQAAINTVAANOE 180  
Db 122 LDAVLADSVDAISRQAVSTIVSANNDLMOQKEDKQAVVDKKAENEKKVKQLEATEA 181  
QY 181 TTAQNTNALNTQQAQLAEALQNLQAEELTTAQQOKATLVAQKAAAE-----E 226





```
Query Match      21.4%; Score 452.5; DB 2; Length 422;
Best Local Similarity 30.0%; Pred. No. 1e-26;
Matches 128; Conservative 81; Mismatches 166; Indels 51; Gaps 12;

QY 28 DPFDAIQSDSKINNLTAAQQAQAAQVNTIQGVYSAQTQQAELQAEQNRLEAQSATLG 87
DB 1 DNVDKKIEEKQEIISLKAKQGDLSQVSSLEAEVSSVFDSEWALREQKOTLLKAKSEQLQ 60

QY 88 QOIQTLSKIVARNESLKQQAARSQAQSNAAATSVINAIINSKVSVDAINRVSAIREVVSA 147
DB 61 QBITNLNQRIEKNEAIKNQARDVQVNGQSTTMDLDAVLDAVDADSAISRQAVSTIVSAN 120

QY 148 EXMLQQEQEDKAAVEQKQEQENQAAINTVAANQETIAQNTNALNTQQAELAAQLNLQABL 207
DB 121 NDLMQQEKEDQAVVDVKKAEKQVQLEATEAELETKRQDLLSKQSELNVKASLALQ 180

QY 208 TTAQDQKATLVAQKAAAE-----EAAQRAAAQAAAEKAAAEKAAKALQEOA 253
DB 181 SSAESKSGKLEKQKAAAEAEQARLAAEQKAAAEKAKQAAKAPAKAEVK--AEAPVASSST 238

QY 254 AQAQAANNNTQATDASDQAAA-----AADNTQAAQTGDSTDQSAQAQAV-----NNSD 302
DB 239 TEAQAPASSS--ATSSSTQQTETTPSTDNSATENTGSSSEQPVQPTTSDNGNNGG 296

QY 303 QESTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNVGGQWMAAS 362
DB 297 QTGGGTVTPTPTPAPSADPTINALNLVLRQSLG-----LRPVV---WDAGLAASAT 345

QY 363 AAAAGYRVGSTPSAGAVAVWNGDGYGHVAVYTVGVQGGQIQV---QEAANYAGNQSTGYRG 419
DB 346 ARAAQVEAGGIPNDH-----WSRG--DEVIAIMWAPGNSVIMAWYNETNMVMTASGSG-HRD 398

QY 420 W-FNPG 424
DB 399 WEINPG 404

RESULT 12
US-09-489-039A-13565
; Sequence 13565, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: 2709-2004001
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13565
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13565

Query Match      13.4%; Score 283.5; DB 2; Length 469;
Best Local Similarity 28.9%; Pred. No. 1e-13;
Matches 113; Conservative 57; Mismatches 164; Indels 57; Gaps 12;

QY 31 DAQIASQDSKINNLTAAQQAQ-----AAQAVNTIQGVYSAQTQQAELQAEQNR 78
DB 84 DAIMVDPGAVNNYRQQAQQAARSARPAEQREKQAQQAQAEELREKQAQAEQRLKQLEQR 143

QY 79 LEAQSATLGQQTQTLSKIVARNESLKQQAARSQAQSNAAATSVINAIINSKVSVDAINRV 138
DB 144 LQAQEAQAEKEQ-----QKQAEAAAKAAAKAAKADQAKAEQAQEAQAAKAA 190

QY 139 AIREVVSAANEKMLQQEQEDKAAVEQKQEQENQAAINTVA--ANQETIAQNTNALNTQQA 197
DB 191 ABAKAKADAQAKAEQQAQAAKAAADAKKQAAEAQAAAEKAAAEKAAAEKAAAEKAAE 250
```

```
QY 198 AA---QLNLQABLTTTAQDQKATLVAQKAAAEBAARQAAAAQAAAAAEKAAAEKALQEOA 254
DB 251 AAAAKAQQAQEAQKQAQQAQKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAA 309

QY 255 QQAQAANNNTQATDASDQQAQAADNTQAAQTGDSTD-----QSAQAQVNNSDQES 305
DB 310 AAKEAQAAD---KAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 366

QY 306 TTATAQAQPSASSASTA-----AVAANTSSANTYPAGQCTGWGKSLAP--WVGNY 352
DB 367 AAAAGSGNTKNSASGADINNYAGQIKSAIESKFYDASSYAGTKTCTLRK-LAPDGLLLNI 425

QY 353 WNGGQWA-ASAAAAGYRVGSTPSAGAVAVW 382
DB 426 QSEGGDPALCQAALAAARQAQKFPKPPSQAVY 456

RESULT 13
US-09-134-001C-5035
; Sequence 5035, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5035
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5035

Query Match      11.2%; Score 237; DB 2; Length 264;
Best Local Similarity 34.7%; Pred. No. 1.7e-10;
Matches 60; Conservative 27; Mismatches 70; Indels 16; Gaps 5;

QY 260 ANNNNTQATDASDQQAQAADNTQAAQTGDSTDQSAQAQVNNSDQESTTTATAQPSASSAS 319
DB 104 SNYNNVQ-----SNNTQSORTTQPTGGLGASYSSTSSNVHVTTSA-PSSNGVYS 151

QY 320 TAAVAANTSSANTYPAGQCTGWV-KSLAPWVGNVWNGGQWMAASAAAAGYRVGSTPSAGA 378
DB 152 LS--NARSASGNLYTSGQCTYYVFDVRVGGKIGSTWGNANNWANAARSGYTTVNNSPA 209

QY 379 VAVWNDGGYGHVAVYTVGV-QGGQIQVQEAANYAGNQSIGNYRGWFPNGSVSYIY 430
DB 210 ILQTSQAGYGHVAYVEGVNSNGSIRVSENNYGHGAGVVTSTRITISASQAASVNY 262

RESULT 14
US-09-710-279-3244
; Sequence 3244, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3244
; LENGTH: 257
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-3244

Query Match      11.2%; Score 236; DB 2; Length 257;
Best Local Similarity 34.9%; Pred. No. 2e-10;
Matches 60; Conservative 26; Mismatches 70; Indels 16; Gaps 5;

QY 261 NNNNTQATDASDQAAAAADNTQAAQTGDSDDQSAQAQAVNNNSDQESTTATAAQPSSASSAST 320
DB 98 NNNYQ-----SNNTSQRTTQPTGGLGASYSTSSNVHVTTSA-PSSNGVSL 145

QY 321 AAVAANTSANTYPAGQCTWGV-KSLAPWVGNWYNGGOWAASAAAAGYRVGSTPSAGAV 379
DB 146 S--NARSASGNLYTSGQCTYYVFDRVGGKIGTGWGNANNWANAARSGYTVNNSPAKGAI 203

QY 380 AVWNDGGYGHVAYVTGV-QGGOIQVOEANYAGNQSIGNYRGWFNPGSVSYIY 430
DB 204 LQTSQAGHYAVYEGVNSNGSIRVSEMMYGHGAGVVTSTRTISASQAASINY 255

RESULT 15
US-09-134-001C-4539
; Sequence 4539, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4539
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4539

Query Match      11.2%; Score 236; DB 2; Length 267;
Best Local Similarity 34.9%; Pred. No. 2.1e-10;
Matches 60; Conservative 26; Mismatches 70; Indels 16; Gaps 5;

QY 261 NNNNTQATDASDQAAAAADNTQAAQTGDSDDQSAQAQAVNNNSDQESTTATAAQPSSASSAST 320
DB 108 NNNYQ-----SNNTSQRTTQPTGGLGASYSTSSNVHVTTSA-PSSNGVSL 155

QY 321 AAVAANTSANTYPAGQCTWGV-KSLAPWVGNWYNGGOWAASAAAAGYRVGSTPSAGAV 379
DB 156 S--NARSASGNLYTSGQCTYYVFDRVGGKIGTGWGNANNWANAARSGYTVNNSPAKGAI 213

QY 380 AVWNDGGYGHVAYVTGV-QGGOIQVOEANYAGNQSIGNYRGWFNPGSVSYIY 430
DB 214 LQTSQAGHYAVYEGVNSNGSIRVSEMMYGHGAGVVTSTRTISASQAASINY 265

Search completed: February 15, 2006, 18:11:57
Job time : 32.0232 secs
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:29:27 ; Search time 92.071 Seconds  
(without alignments)  
1960.467 Million cell updates/sec

Title: US-10-797-821-31  
Perfect score: 2110  
Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNYRGWNPFGSVIYPN 432

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA\_Main:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description          |
|------------|--------|-------------|--------|-------|----------------------|
| 1          | 2110   | 100.0       | 432    | 4     | US-10-383-930-31     |
| 2          | 2110   | 100.0       | 432    | 5     | US-10-797-821-31     |
| 3          | 2099   | 99.5        | 432    | 5     | US-10-383-930-32     |
| 4          | 2099   | 99.5        | 432    | 5     | US-10-797-821-32     |
| 5          | 2083.5 | 98.7        | 431    | 4     | US-10-383-930-30     |
| 6          | 2083.5 | 98.7        | 431    | 5     | US-10-797-821-30     |
| 7          | 2079.5 | 98.6        | 431    | 4     | US-10-383-930-29     |
| 8          | 2079.5 | 98.6        | 431    | 5     | US-10-797-821-29     |
| 9          | 2075.5 | 98.4        | 431    | 4     | US-10-383-930-33     |
| 10         | 2075.5 | 98.4        | 431    | 5     | US-10-797-821-33     |
| 11         | 1087   | 51.5        | 398    | 5     | US-10-474-792-600    |
| 12         | 1059   | 50.2        | 392    | 5     | US-10-472-928-4652   |
| 13         | 1059   | 50.2        | 399    | 5     | US-10-617-320-3230   |
| 14         | 588.5  | 27.9        | 524    | 4     | US-10-282-122A-57658 |
| 15         | 561    | 26.6        | 210    | 4     | US-10-154-251-67     |
| 16         | 500.5  | 23.7        | 449    | 3     | US-09-071-035-482    |
| 17         | 500.5  | 23.7        | 449    | 4     | US-10-206-576-482    |
| 18         | 500.5  | 23.7        | 449    | 5     | US-10-912-362-482    |
| 19         | 452.5  | 21.4        | 422    | 3     | US-09-071-035-484    |
| 20         | 452.5  | 21.4        | 422    | 4     | US-10-206-576-484    |
| 21         | 452.5  | 21.4        | 422    | 5     | US-10-912-362-484    |
| 22         | 325    | 15.4        | 630    | 5     | US-10-494-674-6      |
| 23         | 319    | 15.1        | 600    | 3     | US-09-738-626-5197   |
| 24         | 302.5  | 14.3        | 609    | 5     | US-10-501-282-3184   |
| 25         | 295.5  | 14.0        | 440    | 5     | US-10-501-282-3056   |
| 26         | 294    | 13.9        | 422    | 5     | US-10-501-282-3054   |
| 27         | 285    | 13.5        | 377    | 5     | US-10-501-282-3052   |

|    |       |      |      |   |                      |                   |
|----|-------|------|------|---|----------------------|-------------------|
| 28 | 258.5 | 12.3 | 428  | 4 | US-10-282-122A-55748 | Sequence 55748, A |
| 29 | 252   | 11.9 | 558  | 5 | US-10-732-923-3295   | Sequence 3295, A  |
| 30 | 243   | 11.5 | 484  | 4 | US-10-282-122A-50526 | Sequence 50526, A |
| 31 | 240.5 | 11.4 | 359  | 5 | US-10-650-274-86     | Sequence 86, Appl |
| 32 | 238.5 | 11.3 | 1545 | 5 | US-10-732-923-2248   | Sequence 2248, Ap |
| 33 | 237   | 11.2 | 264  | 4 | US-10-724-972A-6539  | Sequence 6539, Ap |
| 34 | 236   | 11.2 | 257  | 5 | US-10-470-048B-362   | Sequence 362, App |
| 35 | 236   | 11.2 | 267  | 4 | US-10-724-972A-5110  | Sequence 5110, Ap |
| 36 | 233   | 11.0 | 421  | 4 | US-10-282-122A-56483 | Sequence 56483, A |
| 37 | 232   | 11.0 | 2310 | 3 | US-09-874-923-120    | Sequence 120, App |
| 38 | 232   | 11.0 | 2310 | 3 | US-09-991-496-120    | Sequence 120, App |
| 39 | 232   | 11.0 | 2310 | 3 | US-09-820-843A-114   | Sequence 114, App |
| 40 | 231.5 | 11.0 | 2354 | 3 | US-09-820-843A-113   | Sequence 113, App |
| 41 | 230.5 | 10.9 | 323  | 4 | US-10-282-122A-59321 | Sequence 59321, A |
| 42 | 229   | 10.9 | 1463 | 3 | US-09-971-536-69     | Sequence 69, Appl |
| 43 | 226.5 | 10.7 | 261  | 5 | US-10-470-048B-65    | Sequence 65, Appl |
| 44 | 226.5 | 10.7 | 265  | 5 | US-10-470-048B-89    | Sequence 89, Appl |
| 45 | 226.5 | 10.7 | 267  | 5 | US-10-470-048B-74    | Sequence 74, Appl |

ALIGNMENTS

RESULT 1

US-10-383-930-31  
; Sequence 31, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 31  
; LENGTH: 432  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-31

|                       |                 |  |           |             |
|-----------------------|-----------------|--|-----------|-------------|
| Query Match           | 100.0%          | Score 2110;  | DB 4;     | Length 432; |
| Best Local Similarity | 100.0%          | Pred. No. 3.1e-119;  |           |             |
| Matches 432;          | Conservative 0; | Mismatches 0;  | Indels 0; | Gaps 0;     |
| QY                    | 1               | MKKRILSAVLVSGVTLSATLSAIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG    | 60        |             |
| DB                    | 1               | MKKRILSAVLVSGVTLSATLSAIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG    | 60        |             |
| QY                    | 61              | QVSALQTOQAELOAENRLEAQSATLGGQIQTLSSKIVARNESLKQOARSQAQSNAAATSY | 120       |             |
| DB                    | 61              | QVSALQTOQAELOAENRLEAQSATLGGQIQTLSSKIVARNESLKQOARSQAQSNAAATSY | 120       |             |
| QY                    | 121             | INAIINNSVSDAINRVSAIREVVSANEKMLQOQDAAVSKQOENQAANTVAANOE       | 180       |             |
| DB                    | 121             | INAIINNSVSDAINRVSAIREVVSANEKMLQOQDAAVSKQOENQAANTVAANOE       | 180       |             |
| QY                    | 181             | TTAONTNALTQQAQLEAAQLNLQALTTAODQKATLVAQKAAAEBAARQAQAAAEAA     | 240       |             |
| DB                    | 181             | TTAONTNALTQQAQLEAAQLNLQALTTAODQKATLVAQKAAAEBAARQAQAAAEAA     | 240       |             |
| QY                    | 241             | KAAAEKALQEQAQAQAQAANNNNTQATDASDQQAQAANTQAQTGSDTDSQAQAVNN     | 300       |             |
| DB                    | 241             | KAAAEKALQEQAQAQAQAANNNNTQATDASDQQAQAANTQAQTGSDTDSQAQAVNN     | 300       |             |
| QY                    | 301             | SDQESTTATAOPSPASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWGNVWNGGQWA   | 360       |             |
| DB                    | 301             | SDQESTTATAOPSPASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWGNVWNGGQWA   | 360       |             |

QY 361 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANTAGNOSIGNYRGW 420  
DB |||||||  
QY 361 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANTAGNOSIGNYRGW 420  
DB |||||||  
QY 421 FNPGSVSIYPN 432  
DB |||||||  
QY 421 FNPGSVSIYPN 432  
DB |||||||

RESULT 2  
US-10-797-821-31  
; Sequence 31, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 31  
; LENGTH: 432  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-821-31

Query Match 100.0%; Score 2110; DB 5; Length 432;  
Best Local Similarity 100.0%; Pred. No. 3.1e-119;  
Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSATLSA KADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIG 60  
DB |||||||  
QY 1 MKKRILSAVLVSGVTLSATLSA KADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIG 60  
DB |||||||  
QY 61 QVSALOTQAELOAENORLEAQSATLGGQIQTLSSKIARNESLKQARSQAQSNAAATSY 120  
DB |||||||  
QY 61 QVSALOTQAELOAENORLEAQSATLGGQIQTLSSKIARNESLKQARSQAQSNAAATSY 120  
DB |||||||  
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQODKAAVEQKQENQAATNTVAANO 180  
DB |||||||  
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQODKAAVEQKQENQAATNTVAANO 180  
DB |||||||  
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAE LTTAODOKATLVAQKAAAEARQAAAAQAAAE 240  
DB |||||||  
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAE LTTAODOKATLVAQKAAAEARQAAAAQAAAE 240  
DB |||||||  
QY 241 KAAAEAKALQEAQAQAQAANNNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300  
DB |||||||  
QY 241 KAAAEAKALQEAQAQAQAANNNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300  
DB |||||||  
QY 301 SDOESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 360  
DB |||||||  
QY 361 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANTAGNOSIGNYRGW 420  
DB |||||||  
QY 361 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANTAGNOSIGNYRGW 420  
DB |||||||

RESULT 4  
US-10-797-821-32  
; Sequence 32, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens

QY 421 FNPGSVSIYPN 432  
DB |||||||  
QY 421 FNPGSVSIYPN 432  
DB |||||||  
US-10-383-930-32  
; Sequence 32, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 32  
; LENGTH: 432  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-32

Query Match 99.5%; Score 2099; DB 4; Length 432;  
Best Local Similarity 99.3%; Pred. No. 1.4e-118;  
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSATLSA KADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIG 60  
DB |||||||  
QY 1 MKKRILSAVLVSGVTLSATLSA KADDFDAQIASQDSKINNLTAAQQAAQAQVNTTIG 60  
DB |||||||  
QY 61 QVSALOTQAELOAENORLEAQSATLGGQIQTLSSKIARNESLKQARSQAQSNAAATSY 120  
DB |||||||  
QY 61 QVSALOTQAELOAENORLEAQSATLGGQIQTLSSKIARNESLKQARSQAQSNAAATSY 120  
DB |||||||  
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQODKAAVEQKQENQAATNTVAANO 180  
DB |||||||  
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQODKAAVEQKQENQAATNTVAANO 180  
DB |||||||  
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAE LTTAODOKATLVAQKAAAEARQAAAAQAAAE 240  
DB |||||||  
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAE LTTAODOKATLVAQKAAAEARQAAAAQAAAE 240  
DB |||||||  
QY 241 KAAAEAKALQEAQAQAQAANNNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300  
DB |||||||  
QY 241 KAAAEAKALQEAQAQAQAANNNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300  
DB |||||||  
QY 301 SDOESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 360  
DB |||||||  
QY 301 SDOESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA 360  
DB |||||||  
QY 361 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANTAGNOSIGNYRGW 420  
DB |||||||  
QY 361 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANTAGNOSIGNYRGW 420  
DB |||||||  
QY 421 FNPGSVSIYPN 432  
DB |||||||  
QY 421 FNPGSVSIYPN 432  
DB |||||||

RESULT 4  
US-10-797-821-32  
; Sequence 32, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens

```

; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 32
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-32

Query Match          99.5%; Score 2099; DB 5; Length 432;
Best Local Similarity 99.3%; Pred. No. 1.4e-118;
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAISKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIIG 60
DB 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIIG 60

QY 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNATSY 120
DB 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNATSY 120

QY 121 INAIINSKSVSDAIRVSAIREVVSANEKMLQOQEQDKAAVEQKQOENQAANTVAANOE 180
DB 121 INAIINSKSVSDAIRVSAIREVVSANEKMLQOQEQDKAAVEQKQOENQAANTVAANOE 180

QY 181 TTAQNTNALNTQQAQLAAQLNLQAEELTTAQDKATLVAQKAAAEAEARQAAAAQAABEA 240
DB 181 TTAQNTNALNTQQAQLAAQLNLQAEELTTAQDKATLVAQKAAAEAEARQAAAAQAABEA 240

QY 241 KAAAEAKALQEQAAQAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAQVNN 300
DB 241 KAAAEAKALQEQAAQAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAQVNN 300

QY 301 SDOESTTATAOPSSASTAAVAANTSSANTYPAGCCTGWGKSLAPWVNTWNGGGQWA 360
DB 301 SDOESTTATAOPSSASTAAVAANTSSANTYPAGCCTGWGKSLAPWVNTWNGGGQWA 360

QY 361 ASAAAAGYRVGSTPSGAGAVAVWNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
DB 361 ASAAAAGYRVGSTPSGAGAVAVWNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420

QY 421 FNPGSVSYIYPN 432
DB 421 FNPGSVSYIYPN 432

RESULT 5
US-10-383-930-30
; Sequence 30, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08

```

```

; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 30
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-30

Query Match          98.7%; Score 2083.5; DB 4; Length 431;
Best Local Similarity 99.1%; Pred. No. 1.2e-117;
Matches 428; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAISKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIIG 60
DB 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIIG 60

QY 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNATSY 120
DB 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNATSY 120

QY 121 INAIINSKSVSDAIRVSAIREVVSANEKMLQOQEQDKAAVEQKQOENQAANTVAANOE 180
DB 121 INAIINSKSVSDAIRVSAIREVVSANEKMLQOQEQDKAAVEQKQOENQAANTVAANOE 180

QY 181 TTAQNTNALNTQQAQLAAQLNLQAEELTTAQDKATLVAQKAAAEAEARQAAAAQAABEA 240
DB 181 TTAQNTNALNTQQAQLAAQLNLQAEELTTAQDKATLVAQKAAAEAEARQAAAAQAABEA 240

QY 241 KAAAEAKALQEQAAQAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAQVNN 300
DB 241 KAAAEAKALQEQAAQAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAQVNN 300

QY 301 SDOESTTATAOPSSASTAAVAANTSSANTYPAGCCTGWGKSLAPWVNTWNGGGQWA 360
DB 301 SDOESTTATAOPSSASTAAVAANTSSANTYPAGCCTGWGKSLAPWVNTWNGGGQWA 360

QY 361 ASAAAAGYRVGSTPSGAGAVAVWNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
DB 361 ASAAAAGYRVGSTPSGAGAVAVWNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420

QY 421 FNPGSVSYIYPN 432
DB 421 FNPGSVSYIYPN 431

RESULT 6
US-10-797-821-30
; Sequence 30, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 30

```

```
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-30

Query Match          98.7%; Score 2083.5; DB 5; Length 431;
Best Local Similarity 99.1%; Pred. No. 1.2e-117;
Matches 428; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAISKADDFDQAIASQDSKINNLTAAQQAAQAQVNTTIG 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDQAIASQDSKINNLTAAQQAAQAQVNTTIG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 120
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 180
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 180
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAAINTVAANOGE 240
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAAINTVAANOGE 240
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAQQOKATLVAQKAAAEAEARQAAAAQA 240
Db 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAQQOKATLVAQKAAAEAEARQAAAAQA 240
QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAOTGSDTDSAAQAVNN 300
Db 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAOTGSDTDSAAQAVNN 300
QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAOTGSDTDSAAQAVNN 360
Db 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAOTGSDTDSAAQAVNN 360
QY 301 SDOESTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGQWA 359
Db 301 SDOESTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGQWA 359
QY 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVYTVGVQGGQIQVQEBANYAGNOSIGNYRGW 420
Db 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVYTVGVQGGQIQVQEBANYAGNOSIGNYRGW 420
QY 421 FNPQSVSYIYPN 432
Db 421 FNPQSVSYIYPN 431

RESULT 8
US-10-797-821-29
; Sequence 29, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-29

Query Match          98.6%; Score 2079.5; DB 5; Length 431;
Best Local Similarity 98.8%; Pred. No. 2.1e-117;
Matches 427; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAISKADDFDQAIASQDSKINNLTAAQQAAQAQVNTTIG 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDQAIASQDSKINNLTAAQQAAQAQVNTTIG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAAINTVAANOGE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAAINTVAANOGE 180

; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-29
; Sequence 29, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-29

Query Match          98.6%; Score 2079.5; DB 4; Length 431;
Best Local Similarity 98.8%; Pred. No. 2.1e-117;
Matches 427; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAISKADDFDQAIASQDSKINNLTAAQQAAQAQVNTTIG 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDQAIASQDSKINNLTAAQQAAQAQVNTTIG 60
```

Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEKQKQENQAINNTVAANOE 180  
QY 181 TIAQNTNALNTQQAQLAEALQNLQAEITTAQDQKATLVAKKAAAEAAARQAAAAQAAEA 240  
Db 181 TIAQNTNALNTQQAQLAEALQNLQAEITTAQDQKATLVAKKAAAEAAARQAAAAQAAEA 240  
QY 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSFGSAQAQAVNN 300  
Db 241 KAAAEAKALQEAQAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSFGSAQAQAVNN 299  
QY 301 SDOESTTATAOPSPASSASTAAVAANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 360  
Db 300 SDOESTTATAOPSPASSASTAAVAANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 359  
QY 361 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 420  
Db 360 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 419  
QY 421 FNPGSVSYIYPN 432  
Db 420 FNPGSVSYIYPN 431

RESULT 9  
US-10-383-930-33  
; Sequence 33, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 33  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-33

Query Match 98.4%; Score 2075.5; DB 4; Length 431;  
Best Local Similarity 98.6%; Pred. No. 3.7e-117;  
Matches 426; Conservative 2; Mismatches 3; Indels 1; Gaps 1;  
QY 1 MKKRILSAVLVSGVTLSSTLSSATLTSALKADDFDAQIASQDSKINNLTAAQQAAQAQAVNTIOG 60  
Db 1 MKKRILSAVLVSGVTLSSTLSSATLTSALKADDFDAQIASQDSKINNLTAAQQAAQAQAVNTIOG 60  
QY 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120  
Db 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120  
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEKQKQENQAINNTVAANOE 180  
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEKQKQENQAINNTVAANOE 180  
QY 181 TIAQNTNALNTQQAQLAEALQNLQAEITTAQDQKATLVAKKAAAEAAARQAAAAQAAEA 240  
Db 181 TIAQNTNALNTQQAQLAEALQNLQAEITTAQDQKATLVAKKAAAEAAARQAAAAQAAEA 240  
QY 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSFGSAQAQAVNN 300  
Db 241 KAAAEAKALQEAQAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSFGSAQAQAVNN 299  
QY 301 SDOESTTATAOPSPASSASTAAVAANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 360  
Db 300 SDOESTTATAOPSPASSASTAAVAANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 359  
QY 361 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 420  
Db 360 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 419

Db 300 SDOESTTATEAOPSPASSASTAVVTANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 359  
QY 361 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 420  
Db 360 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 419  
QY 421 FNPGSVSYIYPN 432  
Db 420 FNPGSVSYIYPN 431

RESULT 10  
US-10-797-821-33  
; Sequence 33, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 33  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-821-33

Query Match 98.4%; Score 2075.5; DB 5; Length 431;  
Best Local Similarity 98.6%; Pred. No. 3.7e-117;  
Matches 426; Conservative 2; Mismatches 3; Indels 1; Gaps 1;  
QY 1 MKKRILSAVLVSGVTLSSTLSSATLTSALKADDFDAQIASQDSKINNLTAAQQAAQAQAVNTIOG 60  
Db 1 MKKRILSAVLVSGVTLSSTLSSATLTSALKADDFDAQIASQDSKINNLTAAQQAAQAQAVNTIOG 60  
QY 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120  
Db 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120  
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEKQKQENQAINNTVAANOE 180  
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEKQKQENQAINNTVAANOE 180  
QY 181 TIAQNTNALNTQQAQLAEALQNLQAEITTAQDQKATLVAKKAAAEAAARQAAAAQAAEA 240  
Db 181 TIAQNTNALNTQQAQLAEALQNLQAEITTAQDQKATLVAKKAAAEAAARQAAAAQAAEA 240  
QY 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSFGSAQAQAVNN 300  
Db 241 KAAAEAKALQEAQAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSFGSAQAQAVNN 299  
QY 301 SDOESTTATAOPSPASSASTAAVAANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 360  
Db 300 SDOESTTATEAOPSPASSASTAVVTANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 359  
QY 361 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 420  
Db 360 ASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 419









GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:30:27 ; Search time 8.33977 Seconds  
(without alignments)  
736.166 Million cell updates/sec

Title: US-10-797-821-31  
Perfect score: 2110  
Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNRYGMFPGSVIYPN 432

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 107799 seqs, 14211699 residues 107799  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA\_New:\*  
1: /cgn2\_6/prodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/prodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /cgn2\_6/prodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
4: /cgn2\_6/prodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
5: /cgn2\_6/prodata/2/pubpaa/US05\_NEW\_PUB.pep.\*  
6: /cgn2\_6/prodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
7: /cgn2\_6/prodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
8: /cgn2\_6/prodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID                   | Description        |
|------------|--------|-------------|--------|----------------------|--------------------|
| 1          | 2079.5 | 98.6        | 431    | 7 US-11-052-554A-210 | Sequence 210, App  |
| 2          | 1093   | 51.8        | 398    | 7 US-11-052-554A-252 | Sequence 252, App  |
| 3          | 436.5  | 20.7        | 211    | 7 US-11-052-554A-352 | Sequence 352, App  |
| 4          | 411    | 19.5        | 544    | 7 US-11-052-554A-358 | Sequence 358, App  |
| 5          | 236    | 11.2        | 257    | 6 US-10-793-626-3244 | Sequence 3244, App |
| 6          | 229.5  | 10.9        | 394    | 7 US-11-052-554A-79  | Sequence 79, Appl  |
| 7          | 229    | 10.9        | 971    | 7 US-11-052-554A-3   | Sequence 3, Appl   |
| 8          | 226    | 10.7        | 1236   | 6 US-10-873-528-109  | Sequence 109, App  |
| 9          | 222.5  | 10.5        | 270    | 6 US-10-485-517-413  | Sequence 413, App  |
| 10         | 214    | 10.1        | 149    | 6 US-10-793-626-1682 | Sequence 1682, App |
| 11         | 214    | 10.1        | 157    | 6 US-10-793-626-2870 | Sequence 2870, App |
| 12         | 191.5  | 9.1         | 1562   | 7 US-11-052-554A-211 | Sequence 211, App  |
| 13         | 191.5  | 9.1         | 2101   | 6 US-10-857-780-23   | Sequence 23, Appl  |
| 14         | 187.5  | 8.9         | 1095   | 6 US-10-793-626-3154 | Sequence 3154, App |
| 15         | 186.5  | 8.8         | 1586   | 6 US-10-821-234-901  | Sequence 901, App  |
| 16         | 186    | 8.8         | 1448   | 6 US-10-485-517-212  | Sequence 212, App  |
| 17         | 184.5  | 8.7         | 758    | 6 US-10-485-517-144  | Sequence 144, App  |
| 18         | 184.5  | 8.7         | 1410   | 6 US-10-878-556A-136 | Sequence 136, App  |
| 19         | 181.5  | 8.6         | 1122   | 6 US-10-467-657-6112 | Sequence 6112, App |
| 20         | 181    | 8.6         | 330    | 6 US-10-485-517-415  | Sequence 415, App  |
| 21         | 178.5  | 8.5         | 1565   | 6 US-10-467-657-2704 | Sequence 2704, App |
| 22         | 178    | 8.4         | 1107   | 6 US-10-485-517-145  | Sequence 145, App  |
| 23         | 178    | 8.4         | 1290   | 6 US-10-485-517-141  | Sequence 141, App  |
| 24         | 175    | 8.3         | 1126   | 6 US-10-485-517-248  | Sequence 248, App  |
| 25         | 174    | 8.2         | 834    | 7 US-11-052-554A-212 | Sequence 212, App  |

ALIGNMENTS

RESULT 1  
US-11-052-554A-210  
; Sequence 210, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; PRIOR FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 210  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans UA159  
US-11-052-554A-210

|                       |       |                    |  |                            |        |        |   |      |   |
|-----------------------|-------|--------------------|--|----------------------------|--------|--------|---|------|---|
| Query Match           | 98.6% | Score              | 2079.5   | DB 7                       | Length | 431    |   |      |   |
| Best Local Similarity | 98.8% | Pred. No.          | 2.3e-112                                       |                            |        |        |   |      |   |
| Matches               | 427   | Conservative       | 2  | Mismatches                 | 2      | Indels | 1 | Gaps | 1 |
| Qy                    | 1     | MKKRIISAVLVSGVTLS  | SATLTLSAIKADDFDAQI                             | ASQDSKINNLTAAQQAAQAQVNTIOG | 60     |        |   |      |   |
| Db                    | 1     | MKKRIISAVLVSGVTLS  | SATLTSAVKADDFDAQI                              | ASQDSKINNLTAAQQAAQAQVNTIOG | 60     |        |   |      |   |
| Qy                    | 61    | QVSALOTQQAELQAEQNR | LEAQSAITLGGQITLSSKIVARNESLKQAARSQAQKSNATSY     | 120                        |        |        |   |      |   |
| Db                    | 61    | QVSALOTQQAELQAEQNR | LEAQSAITLGGQITLSSKIVARNESLKQAARSQAQKSNATSY     | 120                        |        |        |   |      |   |
| Qy                    | 121   | INAIINSKSVSDAINRV  | SAIREVVSAANEKMLQQQEQDKAAVEQKQENQAANTVAANQE     | 180                        |        |        |   |      |   |
| Db                    | 121   | INAIINSKSVSDAINRV  | SAIREVVSAANEKMLQQQEQDKAAVEQKQENQAANTVAANQE     | 180                        |        |        |   |      |   |
| Qy                    | 181   | TTIAQNTVALNTQQAQLE | AAQNLNLQAEITTTAQDQKATLVAQKAAAEAAARQAAAAQAAAAEA | 240                        |        |        |   |      |   |
| Db                    | 181   | TTIAQNTVALNTQQAQLE | AAQNLNLQAEITTTAQDQKATLVAQKAAAEAAARQAAAAQAAAAEA | 240                        |        |        |   |      |   |
| Qy                    | 241   | KAAAEAKALQEAQAQAQA | QAQAANNNTQATDASDQQAADANTTQAQACTGSDTDSQAAQAVNN  | 300                        |        |        |   |      |   |
| Db                    | 241   | KAAAEAKALQEAQAQAQA | QAQAANNNTQATDVSQQAADANTTQAQACTGSDTDSQAAQAVNN   | 299                        |        |        |   |      |   |
| Qy                    | 301   | SDQESTTTATAAQPSASS | ASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWA      | 360                        |        |        |   |      |   |

Db 300 SDOESTTATEAPSPASSASTRAVAANTSSANTYPAGCCTGWGKSLAPWGVNWTGNGGQWA 359  
QY 361 ASAAAAGYRVGTPSAGAVVNDGGYGHVAVTVTGQGGQIQVQEBANYAGNOSIGNYRGW 420  
Db 360 ASAAAAGYRVGTPSAGAVVNDGGYGHVAVTVTGQGGQIQVQEBANYAGNOSIGNYRGW 419  
QY 421 FNPQSVSYIYPN 432  
Db 420 FNPQSVSYIYPN 431

RESULT 2  
US-11-052-554A-252  
; Sequence 252, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; PRIOR FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 252  
; LENGTH: 398  
; TYPE: PRT  
; ORGANISM: Streptococcus pyogenes MGAS8232  
US-11-052-554A-252

Query Match 51.8%; Score 1093; DB 7; Length 398;  
Best Local Similarity 53.8%; Pred. No. 5e-56;  
Matches 233; Conservative 58; Mismatches 106; Indels 36; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAIGKADDFDQAIASQDSKINNLTAAQQAQAQVNTTQG 60  
Db 1 MKKRILSAVLVSGVTLSSTATTLSAIGKADDFDQAIASQDSKINNLTAAQQAQAQVNTTQG 57  
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQARSACKSNAATSY 120  
Db 58 QVSSIQSEODKUTARNTLEALSKRPEQEIKAITSQIVARNELKKNQARSATYKNETSGY 117  
QY 121 INAIINSKSVDAINRVSAIREVWSANERKMLQQEQDKAAVEQKQEQENQAANTVVAANQE 180  
Db 118 INALLNKSISDVNRLVAINRAVSANAKLLSQKQKADKVSLEKQAAQNTAINTIAANWA 177  
QY 181 TTAQNTNALNTQQAOLEAAQLNLQAEFLTAAQKQATLVQAQKAAABEAAQAAAAQAAAEA 240  
Db 178 MAEENQNTLRTQANLEATANTLALQLASATEDKANLVAQKEAAEKAALAEALQQAQAKV 237  
QY 241 KAAAEKALQEAQAQAQAANNNNTQATDASDQQAADNTQAAOTGSDTQDQSAQAQAVNN 300  
Db 238 KA-----QEQA-----QQAASVEAKSAITPAPQATPAAQSSNA 272  
QY 301 SDOESTTATAOPSPASSASTAAVAANTSSANTYPAGCCTGWGKSLAPWGVNWTGNGGQWA 360  
Db 273 LEPAALTAPAA-PSARPQT-----SYDSNTYTPVGQCTGWGAKSLAPWAGNWNNGGQWA 325  
QY 361 ASAAAAGYRVGTPSAGAVVNDGGYGHVAVTVTGQ--GGQIQVQEBANYAGNOSTGNVTRG 419  
Db 326 YSAQAAGYRTGTPMVGAIVVNDGGYGHVAVTVTGQGGQIQVQEBANYAGNOSTGNVTRG 385  
QY 420 FNPQSVSYIYPN 432  
Db 386 FNPQSVSYIYPN 398

RESULT 3

US-11-052-554A-352  
; Sequence 352, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; PRIOR FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 352  
; LENGTH: 211  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans UA159  
US-11-052-554A-352

Query Match 20.7%; Score 436.5; DB 7; Length 211;  
Best Local Similarity 44.2%; Pred. No. 7.9e-19;  
Matches 95; Conservative 25; Mismatches 60; Indels 35; Gaps 7;

QY 247 KALQEQAAQAQAANNNNTQATDASDQQAADNTQAAQTGDS-----TDQSAQAQ 297  
Db 3 KQFLEKAVFTVAA-----TAATVVLGNKWADAD-TYTLQEGDGSFFSVQRYHMDAYELAS 56  
QY 298 VNNSD-----QESTTATAAQAQPSASSA-----STAATAANTSSANTYPAGCCTGWGKVS 344  
Db 57 MNGKDIITSLILPGQTLTVNGSAAPDQAAAAPTDTQTATETNDANANTYPVGQCTGWGKA 116  
QY 345 LAPWGVNWTGNGGQWAASAAAGYRVGTPSAGAVVNDGGYGHVAVTVTG-OGGQIQV 403  
Db 117 VATWAGDWGNGGOWASSASAAQGYTVGNTPAVGSIMCWTGCGYGHVAVTVAVGEDGKVQV 176  
QY 404 QEANYAGNOSIGNYRGWEN-----PGSVSYIYPN 432  
Db 177 LESNYKDOQWVDNVRGWDFDPNNSGTPGVSYSIYPN 211

RESULT 4  
US-11-052-554A-358  
; Sequence 358, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; PRIOR FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 358  
; LENGTH: 544  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans UA159  
US-11-052-554A-358

Query Match 19.5%; Score 411; DB 7; Length 544;  
Best Local Similarity 25.3%; Pred. No. 6.7e-17;  
Matches 143; Conservative 78; Mismatches 186; Indels 158; Gaps 19;

QY 4 RILSAVLVSGVTLSSTATLSAIAKDDFPAQTA-----SDSKINNLTAAQQAQAQVNTTQ 59  
Db 2 RKLKVALFASILGLMLAVSSYTAADTEDNQVTISHYNEQAGTFDYNVQAANGKTIQSID 61





Db 187 STYWSDAKWASNAANDGYQVDNTPSGAIMOSTPGPYGHVAYVERINGDGSILISEMNY 246  
QY 409 AGNQSIGNRY 418  
Db 247 ANGPYNMYR 256

RESULT 10  
US-10-793-626-1682  
; Sequence 1682, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; PRIOR FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1682  
; LENGTH: 149  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-1682

Query Match 10.1%; Score 214; DB 6; Length 149;  
Best Local Similarity 39.7%; Pred. No. 2.8e-06;  
Matches 50; Conservative 19; Mismatches 43; Indels 14; Gaps 5;

QY 287 GDSSTQSAQAQVNNSDQESTTTATAAQPSSASSASTAAVAANTSSANTYPAGQCTGWG-KSL 345  
Db 20 GHTTHADAAE--NNNQOQST-----YNYSTTEVSF-SNSGNLYTSGQCTWYVDYKT 67  
QY 346 APWVGNYWNGGOWAASAAAGYRVGSTPSAGAVAVMNDGGYGHVAYVTGVQG-GQIQVQ 404  
Db 68 GKGISTWGNANSWATAAQAAGFTVNNTPPEGAIMQSSEGAFGHVAFVSVNNDGSITVS 127  
QY 405 EANYAG 410  
Db 128 ENMYDG 133

RESULT 11  
US-10-793-626-2870  
; Sequence 2870, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; PRIOR FILING DATE: 2004-03-04  
; CURRENT APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2870  
; LENGTH: 157  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-2870

Query Match 10.1%; Score 214; DB 6; Length 157;  
Best Local Similarity 39.7%; Pred. No. 2.9e-06;  
Matches 50; Conservative 19; Mismatches 43; Indels 14; Gaps 5;

QY 287 GDSSTQSAQAQVNNSDQESTTTATAAQPSSASSASTAAVAANTSSANTYPAGQCTGWG-KSL 345  
Db 20 GHTTHADAAE--NNNQOQST-----YNYSTTEVSF-SNSGNLYTSGQCTWYVDYKT 67  
QY 346 APWVGNYWNGGOWAASAAAGYRVGSTPSAGAVAVMNDGGYGHVAYVTGVQG-GQIQVQ 404  
Db 68 GKGISTWGNANSWATAAQAAGFTVNNTPPEGAIMQSSEGAFGHVAFVSVNNDGSITVS 127  
QY 405 EANYAG 410  
Db 128 ENMYDG 133

RESULT 12  
US-11-052-554A-211  
; Sequence 211, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 211  
; LENGTH: 1562  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans UA159  
US-11-052-554A-211

Query Match 9.1%; Score 191.5; DB 7; Length 1562;  
Best Local Similarity 26.3%; Pred. No. 0.00078;  
Matches 87; Conservative 58; Mismatches 115; Indels 71; Gaps 13;

QY 28 DDFDAQIASQDSKINNLTAAQQAAQAQVNTIQGVYVSALQTQQAELQAEQNRLEAQSAT-- 85  
Db 151 DQYKSDVAHEAEVAKIRAKNQATKEQ-----YEKDMAAHKAEVERINANAASK 200  
QY 86 -----LGOQIQTLSS--KIVARNESLKKQARSA-----QKSNAA--TSYINAI-- 125  
Db 201 TAYEAKLAQYQADLAQVKTNAQAAQYQKALAAQYQAEKLRVQEAANAAYDTAVAN 260  
QY 126 NSKSVSDAINRVSAIREVVVSANEKMLQQOEOQKAAVEQKQENQAAINTVAANQETIAQN 185  
Db 261 NAKNT-----ETAAANEETIRKNATAKAEYETKLAQYQAEKRV-----QE 301  
QY 186 TVALNTQQAQLEAAQNLQAEILTAAQDQKATLVQAQAAAEAAARQAAQAAAEAKAAAE 245  
Db 302 ANAAN-----EADYQAKLTAYQTELARVOKANADAKAAYEAFAVANNNAALTAETAE 352  
QY 246 AKALQEOAAQAAA--ANNNTQATDASDQAAAAADNT--QAAQTGDSDDQSAQAQVNNNS 301  
Db 353 NTAIKQRNENAKATYEALKKQYEDLAAVKKANANEADYQAKLTAYQTELARVOKAN-- 410  
QY 302 DQESTTATAQPSASSASTAAVAANTSSANT 332  
Db 411 ----ADAKAAAYEAFAVANNNAALTAETAE--NT 436

RESULT 13  
US-10-857-780-23  
; Sequence 23, Application US/10857780  
; Publication No. US20050272043A1  
; GENERAL INFORMATION:  
; APPLICANT: ROTH, RICHARD B.  
; APPLICANT: BRAUN, ANDREAS







**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 17:57:02 ; Search time 113.754 Seconds  
(without alignments)  
1668.609 Million cell updates/sec

Title: US-10-797-821-32  
Perfect score: 2116  
Sequence: 1 MKRILSLAVLGVLTSSAT.....SIGNVRGWNPGSVYIYPN 432

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21.\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*  
9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description        |
|------------|--------|-------------|--------|-------|--------------------|
| 1          | 2116   | 100.0       | 432    | 7     | ADD93652 Streptoco |
| 2          | 2116   | 100.0       | 432    | 9     | ADx37275 Streptoco |
| 3          | 2099   | 99.2        | 432    | 7     | ADD93651 Streptoco |
| 4          | 2099   | 99.2        | 432    | 9     | ADx37274 Streptoco |
| 5          | 2074.5 | 98.0        | 431    | 7     | ADD93650 Streptoco |
| 6          | 2074.5 | 98.0        | 431    | 9     | ADx37273 Streptoco |
| 7          | 2070.5 | 97.8        | 431    | 9     | ADD93649 Streptoco |
| 8          | 2070.5 | 97.8        | 431    | 9     | ADx37272 Streptoco |
| 9          | 2070.5 | 97.8        | 431    | 9     | AEb91500 Microbial |
| 10         | 2066.5 | 97.7        | 431    | 7     | ADD93653 Streptoco |
| 11         | 2066.5 | 97.7        | 431    | 9     | ADx37276 Streptoco |
| 12         | 1229.5 | 58.1        | 447    | 5     | ABp29684 Streptoco |
| 13         | 1229.5 | 58.1        | 447    | 8     | ADU69524 S agalact |
| 14         | 1229.5 | 58.1        | 447    | 8     | ADV88392 Streptoco |
| 15         | 1229.5 | 58.1        | 447    | 8     | ADV81808 Streptoco |
| 16         | 1229.5 | 58.1        | 447    | 8     | ADV79645 Streptoco |
| 17         | 1087   | 51.4        | 398    | 9     | AEb91542 Microbial |
| 18         | 1081   | 51.1        | 398    | 5     | ABp25919 Streptoco |
| 19         | 1081   | 51.1        | 398    | 8     | ADr83884 S. pyogen |
| 20         | 1071.5 | 50.6        | 395    | 5     | ABp25918 Streptoco |
| 21         | 1053   | 49.8        | 392    | 6     | ABu02747 S. pneumo |
| 22         | 1053   | 49.8        | 392    | 8     | ADr50226 S.pneumon |
| 23         | 1053   | 49.8        | 392    | 8     | ADr50165 S.pneumon |
| 24         | 1053   | 49.8        | 399    | 8     | ADr94595 Novel S.  |

ALIGNMENTS

|          |    |       |      |     |   |          |           |           |
|----------|----|-------|------|-----|---|----------|-----------|-----------|
| RESULT 1 | 25 | 1053  | 49.8 | 399 | 9 | AEA58465 | Aea58465  | Streptoco |
| ADD93652 | 26 | 1052  | 49.7 | 392 | 8 | ADK47859 | Adk47859  | Streptoco |
| ID       | 27 | 1052  | 49.7 | 392 | 8 | ADT50227 | Adt50227  | S.pneumon |
| XX       | 28 | 679.5 | 32.1 | 461 | 2 | AAR14150 | Aar14150  | MSP encod |
| AC       | 29 | 678.5 | 32.1 | 461 | 2 | AAR14530 | Usp45 pro |           |
| XX       | 30 | 658   | 31.1 | 456 | 5 | ABB55584 | Lactococc |           |
| XX       | 31 | 585.5 | 27.7 | 524 | 6 | ABU29734 | Protein e |           |
| DT       | 32 | 581.5 | 27.5 | 525 | 7 | ADC95468 | Adc95468  | E. faeciu |
| XX       | 33 | 555   | 26.2 | 210 | 2 | ADY22579 | Bacterial |           |
| DE       | 34 | 527   | 24.9 | 482 | 9 | ADV16553 | E. faecal |           |
| XX       | 35 | 527   | 24.9 | 497 | 7 | ADH88105 | Adh88105  | Enterococ |
| XX       | 36 | 494.5 | 23.4 | 449 | 2 | AY00250  | Enterococ |           |
| XX       | 37 | 494.5 | 23.4 | 449 | 5 | ABP43469 | E. faecal |           |
| XX       | 38 | 494.5 | 23.4 | 449 | 6 | ABU88497 | E. faecal |           |
| DE       | 39 | 494.5 | 23.4 | 449 | 6 | ABU13748 | Abu13748  | Enterococ |
| XX       | 40 | 494.5 | 23.4 | 449 | 9 | ADV16734 | E. faecal |           |
| XX       | 41 | 494.5 | 23.4 | 449 | 9 | ADY39216 | Novel Ent |           |
| XX       | 42 | 494.5 | 23.4 | 450 | 7 | ADH87829 | Adh87829  | Enterococ |
| XX       | 43 | 445.5 | 21.1 | 422 | 2 | AY00251  | Enterococ |           |
| XX       | 44 | 445.5 | 21.1 | 422 | 5 | ABP43470 | E faecali |           |
| XX       | 45 | 445.5 | 21.1 | 422 | 6 | ABU88498 | Abu88498  | E. faecal |

ADD93652 standard; protein; 432 AA.  
ADD93652;  
29-JAN-2004 (first entry)  
Streptococcus mutans glucan binding protein-B.  
Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.  
Streptococcus mutans.  
WO2003075845-A2.  
18-SEP-2003.  
07-MAR-2003; 2003WO-US006962.  
07-MAR-2002; 2002US-0363209P.  
08-AUG-2002; 2002US-0402483P.  
(FORS-) FORSYTH INST.  
Smith DJ, Taubman MA;  
WPI; 2003-845091/78.  
GENBANK; AY046413.  
Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.  
Claim 5; Page 8; 49pp; English.  
The present sequence is the protein sequence of the glucan binding protein-B (GbpB) of Streptococcus mutans strain 3S1. The sequence includes a number of human leucocyte antigen (HLA)-binding epitopes. The invention provides immunogenic compositions and vaccines for dental caries. The compositions comprise major histocompatibility complex (MHC) class II protein-binding GbpB peptides, especially HLA-binding peptides, covalently linked with peptide subunits of a glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Dieptopic or multiptopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in



XX 07-MAR-2003; 2003WO-US006962.  
XX PF  
XX 07-MAR-2002; 2002US-0363209P.  
XX PR  
XX 08-AUG-2002; 2002US-0402483P.  
XX PR  
XX (FORS-) FORSYTH INST.  
XX PA  
XX Smith DJ, Taubman MA;  
XX WPI: 2003-845091/78.  
XX GENBANK: AY046412.  
XX  
XX Composition useful as vaccines for dental caries comprises a fragment of  
PT a glucan binding protein-B binding to a major histocompatibility complex  
PT class II protein.  
XX  
XX Claim 5; Page 8; 49pp; English.  
XX  
XX The present sequence is the protein sequence of the glucan binding  
CC protein-B (GbpB) of Streptococcus mutans strain 15Jp2. The sequence  
CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
CC invention provides immunogenic compositions and vaccines for dental  
CC caries. The compositions comprise major histocompatibility complex (MHC)  
CC class II protein-binding GbpB peptides, especially HLA-binding peptides,  
CC covalently linked with peptide subunits of a glucosyltransferase. The  
CC compositions are used in a claimed method of eliciting production of an  
CC antibody in a mammal. Dieptopic or multiepitopic polypeptides can be  
CC prepared synthetically or by recombinant DNA technology. Antibodies  
CC raised against MHC class II binding fragments of GbpB can be used in  
CC passive immunisation.  
XX  
XX Sequence 432 AA;

Query Match 99.2%; Score 2099; DB 7; Length 432;  
Best Local Similarity 99.3%; Pred. No. 1.1e-133;  
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MKKRILSAVLVSGVTLSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60  
DB 1 MKKRILSAVLVSGVTLSATTLTSAIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60  
QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQKSNAAATSY 120  
DB 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQKSNAAATSY 120  
QY 121 INAIINSKVSDAINRVSAIREVVSANESKMLQQEQDQKAAVEQKHQENQAATNTVAANQE 180  
DB 121 INAIINSKVSDAINRVSAIREVVSANESKMLQQEQDQKAAVEQKHQENQAATNTVAANQE 180  
QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAODOKATLVAKAAAEAEARQAAAAQAAAAEA 240  
DB 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAODOKATLVAKAAAEAEARQAAAAQAAAAEA 240  
QY 241 KAAAEAKALQEQAQAQAQAANNNNTQATDASDQQAADNTQAAQTGSDTDSQAQAQVNN 300  
DB 241 KAAAEAKALQEQAQAQAQAANNNNTQATDASDQQAADNTQAAQTGSDTDSQAQAQVNN 300  
QY 301 SDESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGYKSLAPWVNTWNGGOWA 360  
DB 301 SDESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGYKSLAPWVNTWNGGOWA 360  
QY 361 ASAAAAGRVGSTPAGAVVNDGGYGHVAVVTGVQGGQIQVQENYAGNQSIGNYRGW 420  
DB 361 ASAAAAGRVGSTPAGAVVNDGGYGHVAVVTGVQGGQIQVQENYAGNQSIGNYRGW 420  
QY 421 FNPGSVSYIPN 432  
DB 421 FNPGSVSYIPN 432

ID ADX37274 standard; protein; 432 AA.  
XX AC  
XX ADX37274;  
XX DT  
XX 21-APR-2005 (first entry)  
XX DE  
XX Streptococcus mutant glucan binding protein B variant #3.  
XX KW immunogenicity; immune stimulation; glucan binding protein-B;  
KW microparticle; major histocompatibility complex; tooth disease.  
XX OS Streptococcus mutans.  
XX PN US2005031633-A1.  
XX PD 10-FEB-2005.  
XX PF 09-MAR-2004; 2004US-00797821.  
XX PR 13-APR-1998; 98US-0081550P.  
XX PR 08-JAN-1999; 99US-0115142P.  
XX PR 12-APR-1999; 99US-00290049.  
XX PR 07-MAR-2002; 2002US-0363209P.  
XX PR 08-AUG-2002; 2002US-0402483P.  
XX PR 07-MAR-2003; 2003US-00383930.  
XX (SMIT/) SMITH D J.  
PA (TAUB/) TAUBMAN M A.  
XX Smith DJ, Taubman MA;  
PI WPI: 2005-151644/16.  
XX  
XX New composition comprising a fragment of a glucan binding protein-B  
PT (GbpB) that binds to MHC class II protein, and a biocompatible  
PT microparticle, useful for producing an antibody (claimed) for immunizing  
PT mammals against dental caries.  
XX  
XX Claim 3; SEQ ID NO 31; 73pp; English.  
XX  
XX The invention relates to a composition comprising a fragment of a glucan  
CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
CC fragment binds to a major histocompatibility complex (MHC) class II  
CC protein. The composition is useful for producing an antibody for  
CC immunizing mammals against dental caries. This sequence corresponds to a  
CC Streptococcus mutans GbpB protein of the invention.  
XX  
XX Sequence 432 AA;  
QY 1 MKKRILSAVLVSGVTLSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60  
DB 1 MKKRILSAVLVSGVTLSATTLTSAIKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60  
QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQKSNAAATSY 120  
DB 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQKSNAAATSY 120  
QY 121 INAIINSKVSDAINRVSAIREVVSANESKMLHQEQDQKAAVEQKHQENQAATNTVAANQE 180  
DB 121 INAIINSKVSDAINRVSAIREVVSANESKMLHQEQDQKAAVEQKHQENQAATNTVAANQE 180  
QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAODOKATLVAKAAAEAEARQAAAAQAAAAEA 240  
DB 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAODOKATLVAKAAAEAEARQAAAAQAAAAEA 240  
QY 241 KAAAEAKALQEQAQAQAQAANNNNTQATDASDQQAADNTQAAQTGSDTDSQAQAQVNN 300  
DB 241 KAAAEAKALQEQAQAQAQAANNNNTQATDASDQQAADNTQAAQTGSDTDSQAQAQVNN 300



CC protein. The composition is useful for producing an antibody for  
CC immunizing mammals against dental caries. This sequence corresponds to a  
CC Streptococcus mutans GbpB protein of the invention.  
XX  
SQ Sequence 431 AA;  
  
Query Match 98.0%; Score 2074.5; DB 9; Length 431;  
Best Local Similarity 98.8%; Pred. No. 5.1e-132;  
Matches 427; Conservative 1; Mismatches 3; Indels 1; Gaps 1;  
  
QY 1 MKKRILSAVLVSGVTLSSATTLSSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60  
DB 1 MKKRILSAVLVSGVTLSSATTLSSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60  
  
QY 61 QVSALQTOQAEQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNSAATSY 120  
DB 61 QVSALQTOQAEQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNSAATSY 120  
  
QY 121 INAINSKSVDAINRVSAIREVVSANEXMLHQOEQDKAAVEQKHQENQAANTVAANOE 180  
DB 121 INAINSKSVDAINRVSAIREVVSANEXMLHQOEQDKAAVEQKHQENQAANTVAANOE 180  
  
QY 181 TTAQNTNALNTQQAQLAQLNLQAEELTTAQDQKATLVAKAAAEAAARQAAAAQAAAA 240  
DB 181 TTAQNTNALNTQQAQLAQLNLQAEELTTAQDQKATLVAKAAAEAAARQAAAAQAAAA 240  
  
QY 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300  
DB 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAVNN 300  
  
QY 301 SDOESTTATAAPSASASASTAAVAANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 360  
DB 301 SDOESTTATAAPSASASASTAAVAANTSSANTYPAGCCTGWKSLAPWVGNVWNGGQWA 360  
  
QY 361 ASAAAAGYVGSPSAGAVVNDGGYGVHVVYTVGGGQIQVQEAANYAGNOSIGNYRGW 420  
DB 361 ASAAAAGYVGSPSAGAVVNDGGYGVHVVYTVGGGQIQVQEAANYAGNOSIGNYRGW 420  
  
QY 421 FNPGSVSYIYPN 432  
DB 420 FNPGSVSYIYPN 431  
  
RESULT 7  
ADD93649  
ID ADD93649 standard; protein; 431 AA.  
XX  
AC ADD93649;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Streptococcus mutans glucan binding protein-B.  
XX  
KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.  
XX  
OS Streptococcus mutans.  
XX  
FH Key Location/Qualifiers  
FT Region 6..25 "HLA-binding peptide"  
FT Region 16..35 "HLA-binding peptide"  
FT Region 33..52 "HLA-binding peptide"  
FT Region 37..56 "HLA-binding peptide"  
FT Region 48..67 "HLA-binding peptide"  
FT Region 52..71 "HLA-binding peptide"  
FT Region 88..107 "HLA-binding peptide"  
FT Region 113..132 "HLA-binding peptide"

FT Region /note= "HLA-binding peptide"  
FT 117..136  
FT /note= "HLA-binding peptide"  
FT 137..156  
FT /note= "HLA-binding peptide"  
FT 174..193  
FT /note= "HLA-binding peptide"  
FT 194..213  
FT /note= "HLA-binding peptide"  
FT 214..233  
FT /note= "HLA-binding peptide"  
FT 248..267  
FT /note= "HLA-binding peptide"  
FT 289..308  
FT /note= "HLA-binding peptide"  
FT 306..325  
FT /note= "HLA-binding peptide"  
FT 311..330  
FT /note= "HLA-binding peptide"  
FT 349..368  
FT /note= "HLA-binding peptide"  
FT 365..384  
FT /note= "HLA-binding peptide"  
FT 383..402  
FT /note= "HLA-binding peptide"  
FT 403..422  
FT /note= "HLA-binding peptide"  
FT WO2003075845-A2.  
XX  
XX 18-SEP-2003.  
PD  
XX  
PF 07-MAR-2003; 2003WO-US006962.  
XX  
PR 07-MAR-2002; 2002US-0363209P.  
PR 08-AUG-2002; 2002US-0402483P.  
XX  
PA (FORS-) FORSYTH INST.  
XX  
XX Smith DJ, Taubman MA;  
XX WPI; 2003-845091/78.  
XX GENBANK; AY046410.  
XX  
PT Composition useful as vaccines for dental caries comprises a fragment of  
PT a glucan binding protein-B binding to a major histocompatibility complex  
PT class II protein.  
XX  
PS Claim 6; Page 7; 49pp; English.  
XX  
XX The present sequence is the protein sequence of the glucan binding  
XX protein-B (GbpB) of Streptococcus mutans strain SK32. The sequence  
XX includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
XX invention provides immunogenic compositions and vaccines for dental  
XX caries. The compositions comprise major histocompatibility complex (MHC)  
XX class II protein-binding GbpB peptides, especially HLA-binding peptides,  
XX covalently linked with peptide subunits of a glucosyltransferase. The  
XX compositions are used in a claimed method of eliciting production of an  
XX antibody in a mammal. Diepitopic or multi-epitopic polypeptides can be  
XX prepared synthetically or by recombinant DNA technology. Antibodies  
XX raised against MHC class II binding fragments of GbpB can be used in  
XX passive immunisation.  
XX  
SQ Sequence 431 AA;  
  
Query Match 97.8%; Score 2070.5; DB 7; Length 431;  
Best Local Similarity 98.6%; Pred. No. 9.5e-132;  
Matches 426; Conservative 1; Mismatches 4; Indels 1; Gaps 1;  
  
QY 1 MKKRILSAVLVSGVTLSSATTLSSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60  
DB 1 MKKRILSAVLVSGVTLSSATTLSSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQOARSQKSNAAATSY 120  
Db |||||  
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQOARSQKSNAAATSY 120  
Db |||||  
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAAINTTVAANQE 180  
Db |||||  
QY 181 TTAQNTNALNTQOAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQOARSQKSNAAATSY 240  
Db |||||  
QY 241 KAAAEKALQEQAAQAAQVAA--NNNTQATDASDQAAAAADNTQAAOTGSDTDSQAAAAQVNN 300  
Db |||||  
QY 241 KAAAEKALQEQAAQAAQVAA--NNNTQATDASDQAAAAADNTQAAOTGSDTDSQAAAAQVNN 299  
QY 301 SQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVNTWNGGGOWA 360  
Db |||||  
QY 361 ASAAAAGYRVGSTPSAGAVAVWVNDGGYGHVAVYVTGVQGGQIQVQEBANYAGNQSIGNYRGW 420  
Db |||||  
QY 421 FNPGSVSIYYPN 432  
Db |||||  
420 FNPGSVSIYYPN 431

RESULT 8  
ADX37272  
ID ADX37272 standard; protein; 431 AA.  
XX  
AC ADX37272;  
XX  
DT 21-APR-2005 (first entry)  
XX  
DE Streptococcus mutant glucan binding protein B variant #1.  
XX  
KW immunogenicity; immune stimulation; glucan binding protein-B;  
KW microparticle; major histocompatibility complex; tooth disease.  
XX  
OS Streptococcus mutans.  
XX  
PN US2005031633-A1.  
XX  
PD 10-FEB-2005.  
XX  
PF 09-MAR-2004; 2004US-00797821.  
XX  
PR 13-APR-1998; 98US-0081550P.  
PR 08-JAN-1999; 99US-0115142P.  
PR 12-APR-1999; 99US-00290049.  
PR 07-MAR-2002; 2002US-0363209P.  
PR 08-AUG-2002; 2002US-0402483P.  
PR 07-MAR-2003; 2003US-00383930.  
XX  
PA (SMIT/) SMITH D J.  
PA (TAUB/) TAUBMAN M A.  
PI Smith DJ, Taubman MA;  
XX WPI; 2005-151644/16.  
XX  
XX New composition comprising a fragment of a glucan binding protein-B  
PT (GbpB) that binds to MHC class II protein, and a biocompatible  
PT microparticle, useful for producing an antibody (claimed) for immunizing  
PT mammals against dental caries.  
XX  
XX Claim 3; SEQ ID NO 29; 73pp; English.  
PS  
XX The invention relates to a composition comprising a fragment of a glucan  
CC binding protein-B (GbpB) and a biocompatible microparticle, where the

CC fragment binds to a major histocompatibility complex (MHC) class II  
CC protein. The composition is useful for producing an antibody for  
CC immunizing mammals against dental caries. This sequence corresponds to a  
CC Streptococcus mutans GbpB protein of the invention.  
XX  
SQ Sequence 431 AA;

Query Match 97.8%; Score 2070.5; DB 9; Length 431;  
Best Local Similarity 98.6%; Pred. No. 9.5e-132; Indels 1; Gaps 1;  
Matches 426; Conservative 1; Mismatches 4;  
QY 1 MKKRILSAVLVSGVTLSSATLTSVAKADDFDAQIASODSKINNLTAQOQAAQAVNTIQG 60  
Db |||||  
QY 1 MKKRILSAVLVSGVTLSSATLTSVAKADDFDAQIASODSKINNLTAQOQAAQAVNTIQG 60  
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQOARSQKSNAAATSY 120  
Db |||||  
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQOARSQKSNAAATSY 120  
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAAINTTVAANQE 180  
Db |||||  
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAAINTTVAANQE 180  
QY 181 TTAQNTNALNTQOAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQOARSQKSNAAATSY 240  
Db |||||  
QY 181 TTAQNTNALNTQOAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQOARSQKSNAAATSY 240  
QY 241 KAAAEKALQEQAAQAAQVAA--NNNTQATDASDQAAAAADNTQAAOTGSDTDSQAAAAQVNN 300  
Db |||||  
QY 241 KAAAEKALQEQAAQAAQVAA--NNNTQATDASDQAAAAADNTQAAOTGSDTDSQAAAAQVNN 299  
QY 301 SQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVNTWNGGGOWA 360  
Db |||||  
QY 301 SQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVNTWNGGGOWA 359  
QY 361 ASAAAAGYRVGSTPSAGAVAVWVNDGGYGHVAVYVTGVQGGQIQVQEBANYAGNQSIGNYRGW 420  
Db |||||  
QY 421 FNPGSVSIYYPN 432  
Db |||||  
420 FNPGSVSIYYPN 431

RESULT 9  
AEB91500  
ID AEB91500 standard; protein; 431 AA.  
XX  
AC AEB91500;  
XX  
DT 20-OCT-2005 (first entry)  
XX  
DE Microbial pathogen adhesin protein sequence, SEQ ID NO:210.  
XX  
KW algorithm; adhesin; pharmaceutical; vaccine; drug screening;  
KW bordetella pertussis infection; antibacterial; pneumonia;  
KW antiinflammatory; respiratory-gen.; gastric ulcer; antitumor;  
KW gastrointestinal-gen.; urinary tract infection; antimicrobial; uropathic.  
XX  
OS Streptococcus mutans.  
XX  
PN WO2005076010-A2.  
XX  
PD 18-AUG-2005.  
XX  
XX 07-FEB-2005; 2005WO-IN000037.  
PF  
XX 06-FEB-2004; 2004IN-DE000173.  
PR  
XX 20-JUL-2004; 2004US-0589227P.  
PR  
XX (COUL ) COUNCIL SCI & IND RES SOUTH AFRICA.  
PA  
XX Sachdeva G, Kumar K, Jain P, Brahmachari SK, Ramachandran S;  
PI



XX WPI; 2005-597835/61.  
XX Computational method for identifying adhesin and adhesin like molecules.  
PT comprises computing sequence-based attributes of protein sequences using  
PT neural network software and training an artificial neural network.  
XX  
XX Claim 16; SEQ ID NO 210; 402pp; English.  
XX  
XX The present invention relates to a computational method (M1) for  
CC identifying adhesin and adhesin-like proteins, by computing the sequence-  
CC based attributes of protein sequences using five attribute modules of a  
CC neural network software, training an artificial neural network (ANN) for  
CC each of the computed five attributes, and identifying the adhesin and  
CC adhesin-like proteins having probability of being an adhesin (Pad) as  
CC equal or greater than 0.51. Also claimed is a set of 274 annotated genes  
CC encoding adhesin and adhesin-like proteins, having 274 fully defined 162-  
CC 11151 base pair (SEQ ID NO: 385-658) sequences; a set of 105 hypothetical  
CC genes encoding adhesin and adhesin-like proteins, having 105 fully  
CC defined 306-15876 base pairs (SEQ ID NO: 659-763); a set of 279 annotated  
CC adhesin and adhesin-like proteins, having 279 fully defined 53-3716 base  
CC pair (SEQ ID NO: 1-279); a set of 105 hypothetical adhesin and adhesin-  
CC like proteins, having 105 fully defined 106-5291 base pair (SEQ ID NO:  
CC 280-384) sequences; and a fully connected multilayer feed forward ANN (I)  
CC based on (M1). (M1) is useful for identifying adhesin and adhesin-like  
CC proteins, of therapeutic potential, and identifying and short-listing  
CC proteins for further testing in development of new vaccine formulations  
CC to eliminate diseases caused by various pathogenic organisms. (M1) is  
CC useful for identifying putative adhesins that are important in drug  
CC discovery and preventing therapeutics for whooping cough, pneumonia, from  
CC gastric ulcer and urinary tract infections. (M1) identifies adhesins from  
CC distantly related organisms, and from bacteria belonging to a wide  
CC phylogenetic spectrum. (M1) is capable of predicting adhesive nature of  
CC unique proteins. The present sequence is a microbial pathogen adhesin  
CC protein sequence.  
XX  
XX Sequence 431 AA;  
XX  
XX Query Match 97.8%; Score 2070.5; DB 9; Length 431;  
XX Best Local Similarity 98.6%; Pred. No. 9.5e-132;  
XX Matches 426; Conservative 1; Mismatches 4; Indels 1; Gaps 1;  
QY 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60  
DB 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60  
QY 61 QVSALQTOQAELOAENRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNSAATSY 120  
DB 61 QVSALQTOQAELOAENRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNSAATSY 120  
QY 121 INAIINSKSVDAINRVSAIREVVSANERKMLHQEQDQKAAVEQKHQENQAANTVAANO 180  
DB 121 INAIINSKSVDAINRVSAIREVVSANERKMLHQEQDQKAAVEQKHQENQAANTVAANO 180  
QY 181 TTAQNTNALNTQQAQLEAQAQLNLQAEELTTAQQOKATLVAKKAAAEAAARQAQAAAA 240  
DB 181 TTAQNTNALNTQQAQLEAQAQLNLQAEELTTAQQOKATLVAKKAAAEAAARQAQAAAA 240  
QY 241 KAAAEKALQEAQAQAQAANNNTQATDASDQQAADNTQAAQTGSDTDSQAQAVNN 300  
DB 241 KAAAEKALQEAQAQAQAANNNTQATDASDQQAADNTQAAQTGSDTDSQAQAVNN 299  
QY 301 SQEESTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 360  
DB 300 SQEESTTATEAPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 359  
QY 361 ASAAAAGYRVGTPSPSAGAVVNDGGYGHVAVYTVGGQGIQVQEANYAGNOSIGNYRGW 420  
DB 360 ASAAAAGYRVGTPSPSAGAVVNDGGYGHVAVYTVGGQGIQVQEANYAGNOSIGNYRGW 419  
QY 421 FNPGSVSIYYPN 432  
DB 420 FNPGSVSIYYPN 431

RESULT 10  
ADD93653  
ID ADD93653 standard; protein; 431 AA.  
XX  
XX ADD93653;  
AC  
XX  
XX 29-JAN-2004 (first entry)  
DT  
XX  
XX Streptococcus mutans glucan binding protein-B.  
DE  
XX  
XX Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.  
KW  
XX  
XX Streptococcus mutans.  
OS  
XX  
XX WO2003075845-A2.  
PN  
XX  
XX 18-SEP-2003.  
PD  
XX  
XX 07-MAR-2003; 2003WO-US006962.  
PF  
XX  
XX 07-MAR-2002; 2002US-0363209P.  
PR  
XX  
XX 08-AUG-2002; 2002US-0402483P.  
PR  
XX  
XX (FORS-) FORSYTH INST.  
PA  
XX  
XX Smith DJ, Taubman MA;  
PI  
XX  
XX WPI; 2003-845091/78.  
DR  
XX  
XX GENBANK; AY046414.  
DR  
XX  
XX Composition useful as vaccines for dental caries comprises a fragment of  
PT a glucan binding protein-B binding to a major histocompatibility complex  
PT class II protein.  
PT  
XX  
XX Claim 5; Page 8-9; 49pp; English.  
XX  
XX The present sequence is the protein sequence of the glucan binding  
XX protein-B (GbpB) of Streptococcus mutans strain SSM3. The sequence  
CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
CC invention provides immunogenic compositions and vaccines for dental  
CC caries. The compositions comprise major histocompatibility complex (MHC)  
CC class II protein-binding GbpB peptides, especially HLA-binding peptides,  
CC covalently linked with peptide subunits of a glucosyltransferase. The  
CC compositions are used in a claimed method of eliciting production of an  
CC antibody in a mammal. Dieptopic or multiptopic polypeptides can be  
CC prepared synthetically or by recombinant DNA technology. Antibodies  
CC raised against MHC class II binding fragments of GbpB can be used in  
CC passive immunisation.  
XX  
XX Sequence 431 AA;  
XX  
XX Query Match 97.7%; Score 2066.5; DB 7; Length 431;  
XX Best Local Similarity 98.4%; Pred. No. 1.8e-131;  
XX Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;  
QY 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60  
DB 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60  
QY 61 QVSALQTOQAELOAENRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNSAATSY 120  
DB 61 QVSALQTOQAELOAENRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNSAATSY 120  
QY 121 INAIINSKSVDAINRVSAIREVVSANERKMLHQEQDQKAAVEQKHQENQAANTVAANO 180  
DB 121 INAIINSKSVDAINRVSAIREVVSANERKMLHQEQDQKAAVEQKHQENQAANTVAANO 180  
QY 181 TTAQNTNALNTQQAQLEAQAQLNLQAEELTTAQQOKATLVAKKAAAEAAARQAQAAAA 240  
DB 181 TTAQNTNALNTQQAQLEAQAQLNLQAEELTTAQQOKATLVAKKAAAEAAARQAQAAAA 240

QY 241 KAAAEAKALOEQAQAQAAANNNTQATDASDQQAARADNTQAAQTGDSQSAQAQAVNN 300  
DB 241 KAAAEAKALOEQAQAQAAA - NNNTQATDASDQQAARADNTQAAQTGDSQSAQAQAVNN 299  
QY 301 SQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGGQWA 360  
DB 300 SQESTTTATEAQPSASSASTAVVTANTSSANTYPAGQCTWGVKSLAPWVGNWNGGGQWA 359  
QY 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQEAANYAGNOSIGNYRCW 420  
DB 360 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQEAANYAGNOSIGNYRCW 419  
QY 421 FNPGSVSIYPN 432  
DB 420 FNPGSVSIYPN 431

RESULT 11  
AD37276  
ID AD37276 standard; protein; 431 AA.  
AC AD37276;  
DT 21-APR-2005 (first entry)  
DE Streptococcus mutant glucan binding protein B variant #5.  
XX immunogenicity; immune stimulation; glucan binding protein-B;  
KW microparticle; major histocompatibility complex; tooth disease.  
XX Streptococcus mutans.  
OS  
XX  
PN US2005031633-A1.  
XX  
PD 10-FEB-2005.  
XX  
PF 09-MAR-2004; 2004US-00797821.  
XX  
PR 13-APR-1998; 98US-0081550P.  
PR 08-JAN-1999; 99US-0115142P.  
PR 12-APR-1999; 99US-0029004P.  
PR 07-MAR-2002; 2002US-0363209P.  
PR 08-AUG-2002; 2002US-0402483P.  
PR 07-MAR-2003; 2003US-00383930.  
XX  
XX (SMIT/) SMITH D J.  
PA (TAUB/) TAUBMAN M A.  
PI Smith DJ, Taubman MA;  
XX  
XX WPI; 2005-151644/16.  
DR  
XX

New composition comprising a fragment of a glucan binding protein-B (GbpB) that binds to MHC class II protein, and a biocompatible microparticle, useful for producing an antibody (claimed) for immunizing mammals against dental caries.  
Claim 3; SEQ ID NO 33; 73pp; English.  
XX The invention relates to a composition comprising a fragment of a glucan binding protein-B (GbpB) and a biocompatible microparticle, where the fragment binds to a major histocompatibility complex (MHC) class II protein. The composition is useful for producing an antibody for immunizing mammals against dental caries. This sequence corresponds to a Streptococcus mutans GbpB protein of the invention.  
XX Sequence 431 AA;  
SQ  
Query Match 97.7%; Score 2066.5; DB 9; Length 431;  
Best Local Similarity 98.4%; Pred. No. 1.8e-131;  
Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;  
QY 1 MKKRILSAVLVSGVTLSSTATLSAVKADDFDAQIASQDSKINNLTAAQQAQAVNTIQG 60

DB 1 MKKRILSAVLVSGVTLSSTATLSAVKADDFDAQIASQDSKINNLTAAQQAQAVNTIQG 60  
QY 61 QVSALQTOQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAQSAQKSAQTSY 120  
DB 61 QVSALQTOQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAQSAQKSAQTSY 120  
QY 121 INAIINSKVSQDINRVSAIREVVSANEKMLHOOEODKAAVEQKHOEQAQNTVAANO 180  
DB 121 INAIINSKVSQDINRVSAIREVVSANEKMLHOOEODKAAVEQKHOEQAQNTVAANO 180  
QY 181 TIAQNTNALNTQQAQLEAAQNLQAEELTTAQDQKATLVAQKAAAEAEAAQAAAAA 240  
DB 181 TIAQNTNALNTQQAQLEAAQNLQAEELTTAQDQKATLVAQKAAAEAEAAQAAAAA 240  
QY 241 KAAAEAKALOEQAQAQAAANNNTQATDASDQQAARADNTQAAQTGDSQSAQAQAVNN 300  
DB 241 KAAAEAKALOEQAQAQAAA - NNNTQATDASDQQAARADNTQAAQTGDSQSAQAQAVNN 299  
QY 301 SQESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGGQWA 360  
DB 300 SQESTTTATEAQPSASSASTAVVTANTSSANTYPAGQCTWGVKSLAPWVGNWNGGGQWA 359  
QY 361 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQEAANYAGNOSIGNYRCW 420  
DB 360 ASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQEAANYAGNOSIGNYRCW 419  
QY 421 FNPGSVSIYPN 432  
DB 420 FNPGSVSIYPN 431

RESULT 12  
ABP29684  
ID ABP29684 standard; protein; 447 AA.  
XX  
AC ABP29684;  
XX  
DT 02-JUL-2002 (first entry)  
DE Streptococcus polypeptide SEQ ID NO 8544.  
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
XX Streptococcus agalactiae.  
OS  
XX WO200234771-A2.  
XX  
XX  
XX PD 02-MAY-2002.  
XX  
XX PF 29-OCT-2001; 2001WO-GB004789.  
XX  
XX PR 27-OCT-2000; 2000GB-00026333.  
PR 24-NOV-2000; 2000GB-00028727.  
PR 07-MAR-2001; 2001GB-00005640.  
XX  
XX (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX Telford J, Massignani V, Margarit Y Rosi, Grandi G, Fraser C;  
PI Tettelin H;  
XX WPI; 2002-352536/38.  
DR N-PSDB; ABN70315.  
XX  
XX New Streptococcus protein for the treatment or prevention of infection or disease caused by Streptococcus bacteria, such as meningitis, and for detecting a compound that binds to the protein.  
PS Claim 1; Page 3965; 4525pp; English.  
XX

CC The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and anti-inflammatory  
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins  
XX  
XX Sequence 447 AA;

Query Match 58.1%; Score 1229.5; DB 5; Length 447;  
Best Local Similarity 59.7%; Pred. No. 7.9e-75;  
Matches 276; Conservative 42; Mismatches 99; Indels 45; Gaps 9;  
QY 1 MKKRILSAVLVSGVTLSATTLSSAVKADDFDAQIASQDSKINNLTAAQOAAQVNTTIQ 60  
Db 1 MKKRILSAVLVSGVTLTAAV--TVNADDFDKIAATDSVINTLSGQAAQVNTAIKG 58  
QY 61 QVSALQTOQAEIQAENORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAOKSNAATSY 120  
Db 59 QVGALESQOQSELEAQAQLEAVSQQLGQEIQTLSNKNIVARNESLKQVRSQKGN-LTNY 117  
QY 121 INAINSKSVSDAINRVSAIREVVSANEKMLHQEQDKAAVQKQHQENQAINTVAANO 180  
Db 118 INTILNSKSVSDAVNVVVAIREVVSANEKMLAQEQEADKALEAKQIENQAINTVAA 177  
QY 181 TTAQNTNALNTQOALEAAQLNLQELTAAQOKATLVAKAABEAARQAQAAQAAABA 240  
Db 178 ALENKAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQAABEAARQAQAAABA 237  
QY 241 KAAAEKALQEOAQAQAAQAAANNNTQATDASDQOAAADNTQAAQTGDSTDOQAQAVNN 300  
Db 238 KAQAQAKAQAESVAKAQAQAA--OVESATAPTETVQTPRTETKPSNLTAT 285  
QY 301 SDQ---ESTTATA-----AQSASSASTAA-----VAANTSANTYPAG 336  
Db 286 SSATTVATTATATNEPKVTQPSVVTKAVEAPKAVVSTTPRAVSKPVVRSYDSSNTYP 345  
QY 337 QCTWGVKSLAPVWGNVWNGGQWAAASAAAAGYRVGSTPSAGAVAVW--NDGGYGHVAVT 394  
Db 346 QCTWGAKSMAVWGNVWGNQWGAASARAAGYSVGTTPRVGAVAVMPYDGGYGHVAVT 405  
QY 395 GV-QGGQIQVQEAANYAGNOSIGNYRWENP---GSVSYIYPN 432  
Db 406 SVANNSSIQVMESNYAGNWSIGNYRGSFNPSAGSGVYIYPN 447

RESULT 13  
ADU69524  
ID ADU69524 standard; protein; 447 AA.  
XX  
AC ADU69524;  
XX  
DT 10-FEB-2005 (first entry)  
XX  
DE S agalactiae hyperimmune serum reactive antigen seqid 219.  
XX  
KW immune stimulation; antigen; bacterial surface display;  
KW hyperimmune serum reactive antigen; vaccine; bacterial infection;  
KW antibacterial; infection.  
XX  
OS Streptococcus agalactiae.  
XX

PN W02004099242-A2.  
XX 18-NOV-2004.  
XX 06-MAY-2004; 2004WO-EP004856.  
XX 07-MAY-2003; 2003EP-00450112.  
PR 28-NOV-2003; 2003EP-00450266.  
XX (INTE-) INTERCELL AG.  
XX Meinke A, Nagy E, Hanmer M, Horky M, Kallenda S, Prustomersky S;  
PI WPI; 2004-821662/81.  
XX N-PSDB; ADU69307.  
XX New nucleic acid molecule encoding a hyperimmune serum reactive antigen,  
PT useful for the manufacture of a vaccine against Streptococcus agalactiae  
PT infection.  
XX Claim 14; SEQ ID NO 219; 221pp; English.

CC The invention describes an isolated nucleic acid molecule encoding a  
CC hyperimmune serum reactive antigen or its fragment. Also described are: a  
CC vector comprising the nucleic acid molecule; a host cell comprising the  
CC vector; a hyperimmune serum-reactive antigen comprising a sequence  
CC encoded by the nucleic acid molecule and consisting of e.g., 85, 299, 467  
CC or 812 amino acids; fragments of hyperimmune serum-reactive antigens  
CC consisting of peptides comprising e.g., 76, 134, 221 or 576 amino acids;  
CC a process for producing a Streptococcus agalactiae hyperimmune serum  
CC reactive antigen; a process for producing a cell that expresses a S.  
CC agalactiae hyperimmune serum reactive antigen; a pharmaceutical  
CC composition, especially a vaccine, comprising the hyperimmune serum-  
CC reactive antigen or nucleic acid molecule; an antibody that binds at  
CC least to a selective part of the hyperimmune serum-reactive antigen; a  
CC hybridoma cell line, which produces the antibody; a method for producing  
CC the antibody; an antagonist that binds to the hyperimmune serum-reactive  
CC antigen; a method for identifying an antagonist capable of binding to the  
CC hyperimmune serum-reactive antigen; a method for identifying an  
CC antagonist capable of reducing or inhibiting the interaction activity of  
CC a hyperimmune serum-reactive antigen to its interaction partner; a  
CC process for in vitro diagnosing a disease related to expression of the  
CC hyperimmune serum-reactive antigen; and a process for in vitro diagnosis  
CC of a bacterial infection, especially a S. agalactiae infection. The  
CC hyperimmune serum reactive antigen is useful for isolating, purifying  
CC and/or identifying an interaction partner of the hyperimmune serum  
CC reactive antigen. The hyperimmune serum reactive antigen is useful for  
CC generating a peptide binding to the hyperimmune serum reactive antigen,  
CC where the peptide comprises anticalines, or for the manufacture of a  
CC functional nucleic acid comprising aptamers or Spiegelmers. The nucleic  
CC acid molecule is useful for the manufacture of a functional ribonucleic  
CC acid comprising ribozymes, antisense nucleic acids or siRNA. The nucleic  
CC acid molecule, hyperimmune serum-reactive antigen or antibody is useful  
CC for the manufacture of a vaccine against S. agalactiae infection. This is  
CC the amino acid sequence of a Streptococcus agalactiae hyperimmune serum  
CC reactive antigen.

XX Sequence 447 AA;  
Query Match 58.1%; Score 1229.5; DB 8; Length 447;  
Best Local Similarity 59.7%; Pred. No. 7.9e-75;  
Matches 276; Conservative 42; Mismatches 99; Indels 45; Gaps 9;  
QY 1 MKKRILSAVLVSGVTLSATTLSSAVKADDFDAQIASQDSKINNLTAAQOAAQVNTTIQ 60  
Db 1 MKKRILSAVLVSGVTLTAAV--TVNADDFDKIAATDSVINTLSGQAAQVNTAIKG 58  
QY 61 QVSALQTOQAEIQAENORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAOKSNAATSY 120  
Db 59 QVGALESQOQSELEAQAQLEAVSQQLGQEIQTLSNKNIVARNESLKQVRSQKGN-LTNY 117  
QY 121 INAINSKSVSDAINRVSAIREVVSANEKMLHQEQDKAAVQKQHQENQAINTVAANO 180  
XX

Db 118 INTILNSKSVSDAVNRVVAIREVVSANEKMLAQEADKAALEAKQIENQNAINTVAANKQ 177  
 QY 181 TTAQNTNALNTQQAQLEAAQLNLQELTTAQDQKATLVAQKAAABEAARQAAAAQAAAEA 240  
 Db 178 AIENNKAAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQABEAARKAAEAQAAAEA 237  
 QY 241 KAAAEKALQEQAAQAAQAAANNNNTQATDASDQQAADNTQAAOTGDSTQDQSAQAQVNN 300  
 Db 238 KAQAEEAKQAESVAKAQAAA-----QVESATAPTETVQTPRTEIKPSNLTAT 285  
 QY 301 SDQ---ESTTATA-----AOPSASSASTAA-----VAANTSSANTYPAG 336  
 Db 286 SSATTVATTATATNEPKVTPQSVVTKAVEAPKAVVSTPRAVSKPVVRSYDSSNTYPWG 345  
 QY 337 QCTGWKSLAPWVGNVGNWGQAAASAAAAGYRVGSTPSAGAVW--NDGGYGHVAVVT 394  
 Db 346 QCTWGAKSMASWVGNVGNWGNANQWGSARAAGYSVGTTPRVGAVAVWVPYDGGYGHVAVVT 405  
 QY 395 GV-QGGQIQOQVQANVAGNOSIGNYRGWNP---GSVSYIYPN 432  
 Db 406 SVANNSSIQVMESYAGNMSIGNYRGSEFNPSAGSGVYIYPN 447

RESULT 14

ADV81808  
 ID ADV81808 standard; protein; 447 AA.  
 AC ADV81808;  
 DT 24-FEB-2005 (first entry)  
 DE Streptococcus agalactiae protein sequence, SEQ ID 786.  
 KW Antibacterial; Vaccine; bacterial infection.  
 OS Streptococcus agalactiae.  
 PN FR2824074-A1.  
 PD 31-OCT-2002.  
 PF 26-APR-2001; 2001FR-00005642.  
 PR 26-APR-2001; 2001FR-00005642.  
 PA (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 PI Glaser P, Rusniok C, Chevallier F, Frangeul L, Lalioui L;  
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;  
 WPI; 2004-101891/11.

Genomic nucleotide sequences encoding polypeptides of Streptococcus agalactiae for the development of vaccines, diagnostic tools, DNA chips and identification of therapeutic targets.

Claim 6; SEQ ID NO 786; 2687pp; French.

The present invention relates to novel Streptococcus agalactiae nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II; ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S. agalactiae involved in the synthesis of amino acids, cell membranes, intermediate (central) metabolism, energetic metabolism, fatty acid and phospholipid metabolism, nucleotide metabolism including purines, pyrimidines and/or nucleosides, regulatory functions, replication, transcription, translation, protein transport, adaptation to atypical conditions, sensitivity to medicines and/or analogues, functions related to transposons, biosynthesis of cofactors, prosthetic groups and transporters, cell membrane proteins and cellular machinery. (I) are useful for the detection and/or amplification of nucleic acids. CC Pharmaceutical composition comprising (I) or (II) are useful for treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is

CC equivalent for the present basic patent FR2824074A1. WO200292818A2  
 CC contains 6617 sequence whereas the present patent only contains 2344  
 CC sequences.  
 XX  
 SQ Sequence 447 AA;

Query Match 58.1%; Score 1229.5; DB 8; Length 447;  
 Best Local Similarity 59.7%; Pred No. 7.9e-75;  
 Matches 276; Conservative 42; Mismatches 99; Indels 45; Gaps 9;

QY 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASQDSKINNLTAAQOQAAQAVNTIAG 60  
 Db 1 MKKRILSAVLVSGVTLGTAAV--TVNADDFDSKIAATDSVINTLSGQQAQAQVTAIKG 58  
 QY 61 QVSALQTOQAELOAENQRLQEAQSATLGOQIOTLSSKIIVARNESLKQQAARSQAQNAATSY 120  
 Db 59 QVGALESQCSLEAQAQLEAVSQQLGQEIOTLSNKIIVARNESLKKQVRSQAQGN-LTNY 117  
 QY 121 INAIINSKSVSDAINRVSAIREVVSANEKMLHQEQDAAVVEQKHQENQAAINTVAANOE 180  
 Db 118 INTILNSKSVSDAVNRVVAIREVVSANEKMLAQEADKAALEAKQIENQNAINTVAANKQ 177  
 QY 181 TTAQNTNALNTQQAQLEAAQLNLQELTTAQDQKATLVAQKAAABEAARQAAAAQAAAEA 240  
 Db 178 AIENNKAAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQABEAARKAAEAQAAAEA 237  
 QY 241 KAAAEKALQEQAAQAAQAAANNNNTQATDASDQQAADNTQAAOTGDSTQDQSAQAQVNN 300  
 Db 238 KAQAEEAKQAESVAKAQAAA-----QVESATAPTETVQTPRTEIKPSNLTAT 285  
 QY 301 SDQ---ESTTATA-----AOPSASSASTAA-----VAANTSSANTYPAG 336  
 Db 286 SSATTVATTATATNEPKVTPQSVVTKAVEAPKAVVSTPRAVSKPVVRSYDSSNTYPWG 345  
 QY 337 QCTGWKSLAPWVGNVGNWGQAAASAAAAGYRVGSTPSAGAVW--NDGGYGHVAVVT 394  
 Db 346 QCTWGAKSMASWVGNVGNWGNANQWGSARAAGYSVGTTPRVGAVAVWVPYDGGYGHVAVVT 405  
 QY 395 GV-QGGQIQOQVQANVAGNOSIGNYRGWNP---GSVSYIYPN 432  
 Db 406 SVANNSSIQVMESYAGNMSIGNYRGSEFNPSAGSGVYIYPN 447

RESULT 15

ADV81808  
 ID ADV81808 standard; protein; 447 AA.  
 AC ADV81808;  
 DT 24-FEB-2005 (first entry)  
 DE Streptococcus agalactiae protein, SEQ ID 2949.  
 KW Antibacterial; vaccine; bacterial infection.  
 OS Streptococcus agalactiae.  
 PN WO200292818-A2.  
 PD 21-NOV-2002.  
 PF 26-APR-2002; 2002WO-IB003059.  
 PR 26-APR-2001; 2001FR-00005642.  
 PA (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 PI Glaser P, Rusniok C, Chevallier F, Frangeul L, Lalioui L;  
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
 WPI; 2004-101891/11.

PT Genomic nucleotide sequences encoding polypeptides of Streptococcus  
PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
XX and identification of therapeutic targets.

PS Claim 6; SEQ ID NO 2949; 439pp; French.

XX The present invention relates to novel Streptococcus agalactiae  
CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
CC nucleotide sequences encode polypeptides of S. agalactiae involved in the  
CC synthesis of amino acids, cell membranes, intermediate (central)  
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
CC regulatory functions, replication, transcription, translation, protein  
CC transport, adaptation to atypical conditions, sensitivity to medicines  
CC and/or analogues, functions related to transposons, biosynthesis of  
CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
CC cellular machinery. (I) are useful for the detection and/or amplification  
CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
CC useful for treatment of a bacterial S. agalactiae infection. The complete  
CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
CC present patent is an equivalent for the basic patent FR2824074A1, which  
CC contains only 2344 sequences.

XX Sequence 447 AA;

|                       |     |  |                    |            |             |
|-----------------------|-----|--|--------------------|------------|-------------|
| Query Match           |     | 58.1%;   | Score 1229.5;      | DB 8;      | Length 447; |
| Best Local Similarity |     | 59.7%;   | Pred. No. 7.9e-75; |            |             |
| Matches 276;          |     | Conservative 42;   | Mismatches 99;     | Indels 45; | Gaps 9;     |
| QY                    | 1   | MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG | 60                 |            |             |
| Db                    | 1   | MKKRILSAVLVSGVTLTAAV--TVNADDFDSKIAATDSVINTLSQQAQAQVNTAIKG    | 58                 |            |             |
| QY                    | 61  | QVSALQTQQAELQAEORLEAQSATILGQOITLSSKIVARNESLKKOARSQKSNAAATSY  | 120                |            |             |
| Db                    | 59  | QVGALESQOSELEAQAQLEAVSQSLGQEIQTLSNKIVARNESLKKQVRSQKGN-LTNY   | 117                |            |             |
| QY                    | 121 | INAIINSKVSDAINRVSAIREVSWANEKMLHQEQDKAAVEQKHQENQAAINTVAANOE   | 180                |            |             |
| Db                    | 118 | INTILNSKVSDAVNRVVAIREVSVSANEKMLAQEQEADKALEAKQIENQNAINTVAANKQ | 177                |            |             |
| QY                    | 181 | TTAQNNTALNTQQAQLEAAQLNLQAEELTTAQDOKATLVQAQKAAAEAAQAAAAAEEA   | 240                |            |             |
| Db                    | 178 | AIENNKAAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQAEAAKAAEAQAAEA   | 237                |            |             |
| QY                    | 241 | KAAAEAKAQEAQQAQAQAANNNTQATDASDQQAADNTQAAQTGDSQSAQAQVNN       | 300                |            |             |
| Db                    | 238 | KAAQAEAKAQAESVAKAQAAA-----QVESATAPTETVQTPTETKPSNLTAT         | 285                |            |             |
| QY                    | 301 | SDQ---ESTTATA-----AQPSASSASTAA-----VAANTSSANTYPAG            | 336                |            |             |
| Db                    | 286 | SSATTVATTATTATNEPKVTQPSVTVKAVEAPKAVVSTPRAVSKPVVRSYDSSNTYPNG  | 345                |            |             |
| QY                    | 337 | QCTWGVKSLAPVGNVWNGGQWAAASAAAGYRVGSGTSPSAGAVVM--NDGGYGHVAVYT  | 394                |            |             |
| Db                    | 346 | QCTWGAKSMAWVGNVWGNANQWGSARAAGYSGVTTPRVGAVAVMPYDGGYGHVAVYT    | 405                |            |             |
| QY                    | 395 | GV-QGGQIQVQEAQYAGNQSIGNYRGWNP---GSVSVIYYPN                   | 432                |            |             |
| Db                    | 406 | SVANNSSIQWESNYSVAGNMSIGNYRGWNP---GSVSVIYYPN                  | 447                |            |             |

Search completed: February 15, 2006, 18:03:00  
Job time : 114.754 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:03:21 ; Search time 21.0162 Seconds  
(without alignments)  
1977.789 Million cell updates/sec

Title: US-10-797-821-32  
Perfect score: 2116  
Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNYRGWFGVSVIYVPN 432

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 1053  | 49.8        | 392    | 2 G95258 | secreted 45 kd pro |
| 2          | 1053  | 49.8        | 392    | 2 B98124 | general stress pro |
| 3          | 681.5 | 32.2        | 461    | 2 JN0097 | secreted 45K prote |
| 4          | 658   | 31.1        | 456    | 2 E86903 | hypothetical prote |
| 5          | 537.5 | 25.4        | 507    | 2 S05542 | hypothetical prote |
| 6          | 356   | 16.8        | 398    | 2 AC1763 | peptidoglycan lyti |
| 7          | 343.5 | 16.2        | 401    | 2 A11387 | cell wall-binding  |
| 8          | 319   | 15.1        | 473    | 2 F70031 | conserved hypothet |
| 9          | 259.5 | 12.3        | 581    | 2 E75383 | cell wall-binding  |
| 10         | 256   | 12.1        | 461    | 2 H84099 | cell wall binding  |
| 11         | 254.5 | 12.0        | 436    | 2 AH1387 | cell wall binding  |
| 12         | 251   | 11.9        | 437    | 2 AB1763 | conserved hypothet |
| 13         | 237   | 11.2        | 528    | 2 B75310 | conserved hypothet |
| 14         | 230   | 10.9        | 421    | 2 JV0057 | tola protein - Bac |
| 15         | 228   | 10.8        | 971    | 2 B90835 | probable tail fibe |
| 16         | 228   | 10.8        | 973    | 2 C85693 | probable membrane  |
| 17         | 226.5 | 10.7        | 255    | 2 G90061 | hypothetical prote |
| 18         | 226.5 | 10.7        | 265    | 2 B89837 | hypothetical prote |
| 19         | 226.5 | 10.7        | 267    | 2 F90028 | hypothetical prote |
| 20         | 226.5 | 10.7        | 394    | 2 F90725 | membrane spanning  |
| 21         | 226.5 | 10.7        | 394    | 2 G85576 | membrane spanning  |
| 22         | 225.5 | 10.7        | 166    | 2 G90029 | hypothetical prote |
| 23         | 225.5 | 10.7        | 392    | 2 F96937 | cell wall-associat |
| 24         | 222   | 10.5        | 492    | 2 A28616 | M5 protein precurs |
| 25         | 221.5 | 10.5        | 688    | 2 A83179 | conserved hypothet |
| 26         | 221   | 10.4        | 1122   | 2 G64887 | probable tail fibe |
| 27         | 217   | 10.3        | 1528   | 2 A60338 | surface antigen A  |
| 28         | 217   | 10.3        | 4776   | 2 E95206 | cell wall surface  |
| 29         | 215   | 10.2        | 539    | 2 A28549 | M24 protein precur |

RESULT 1  
G95258  
secreted 45 kd protein [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: G95258  
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide  
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel, I.E.  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: G95258  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-392 <KUR>  
A:Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMY4; UNIPARC:UPI0000051B81; GB:AB005672; I  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP2216

Query Match 49.8%; Score 1053; DB 2; Length 392;  
Best Local Similarity 49.8%; Pred. No. 2.3e-45;  
Matches 217; Conservative 74; Mismatches 93; Indels 52; Gaps 4;

QY 1 MKKRILSAVLVSGVTLSATILSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60  
DB 1 MKKRILASLLSTVMVSVQAVLTAAHAEITDDKIAAQDNKISNLTAAQQAQAQVDQIQE 60  
QY 61 QVSALOTQAAELQAEORLEAQSATLGOOIOTLSKIVARNESLKQQAASQKSNAAATSY 120  
DB 61 QVSALQAEQSNLQAEORLEAQSATLGOOIOTLSKIVARNESLKQQAASQKSNAAATSY 120  
QY 121 INAIINSKSVSDAINRVSAIREVWSANEKMLHQEQSDKAAVEQKHQHOEQAQNTVAANQE 180  
DB 121 INAIINSKSVSDAINRVSAIREVWSANEKMLHQEQSDKAAVEQKHQHOEQAQNTVAANQE 180  
QY 181 TIAQNTNALTQQAQLEAAQNLQAEELTAQDQKATLVAQKAAAEAAEAQAAQAAEA 240  
DB 181 KLADDAQALTTKQAEKAAELSLAAEKATAGEKASLLLEQKAAAEAAEAQAAEAAYKE 240  
QY 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQQAADNDTQAAOTGSDTQDQAQAVNN 300  
DB 241 KRASQQQSV-----LASANTLTAQQAQVSESAAAQVRAKVRPT----- 279  
QY 301 SDQESTTATAAQAQPSASSASTAAVAANTSSANTYPAGQCQTGWKSLAPVGNVWNGGQWA 360  
DB 301 SDQESTTATAAQAQPSASSASTAAVAANTSSANTYPAGQCQTGWKSLAPVGNVWNGGQWA 360  
QY 361 ASAAAAGYRVGSTPSAGAVAVWVNDGGYGHVAVTVTQVG-QGQVQEAQYAGNQSGNYRG 419  
DB 361 ASAAAAGYRVGSTPSAGAVAVWVNDGGYGHVAVTVTQVG-QGQVQEAQYAGNQSGNYRG 419

ALIGNMENTS





Db 1 MKKKIISAILMTSVLSSAAPISGVYADT-NSDIAKQDATISSAOSAKAQAQAQVDSLQS 59  
QY 61 QVSALQTOQAELQAEORLEASATLGGQIOTLSSKIVARNESLKQARSQAOKSNAATSY 120  
Db 60 KVDLSQOKQASTAQIAKIESEAKALNAQIATILNESIAERTKTLEQAARSQAQVNSATNY 119  
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAAINTVAANOQE 180  
Db 120 MDVAVNSKSLTDVIQKVTATATVSSANKQMLEQKEQKELSKQSTVKNTYNQFVLSLQ 179  
QY 181 TTAQNTNALNTQQAELQAELNLOAELTTAODQKATLVAQKAAAEAEARQAAQAARAA 240  
Db 180 SLDSQAQELTSQAELKVATLNYQATATAODKQOSLLDEKAAAEKAAQEAQKQAAYEA 239  
QY 241 KAAAEAKALQEQAAQAAANNNNTQATDASDQQAADANTQAAQTGDSTDQSAQA--AV 298  
Db 240 Q-----QKEAQAQAATAATYKAV---BEATSVSSSSQASQSSSSNTSSNTSS 288  
QY 299 NNSDQESTTATAAQPSSASSASTAAVAANTSS-----ANTYPAG 336  
Db 289 NSSSSSSSSSSSSSSSSSSSGGSGTNGNNAAGTNGTGGSSSGINGSTPIANPYAGG 348  
QY 337 QCT---WG-----VKSLAPWVGNWCGNCQWAAASAAAAG--YRVGSTPAGAVAV-- 381  
Db 349 GCTDYVMQYFAAQGYIRIMP-----GNGGQWATNGPAQGVHLVVGGAAP--GVIASSF 400  
QY 382 -----WMDGGYGHVAYVTGV--QGQIOIQVEANYAGNQSIGNVGFNPGSVSYTYPN 432  
Db 401 SADPFGVYANSPIGHVAIVKSVNSNGTIIKEGGY-GTTWGHGR-TVSASGVTFMLPN 456  
RESULT 5  
S05542  
hypothetical protein, 54K - Enterococcus faecium  
C:Species: Enterococcus faecium  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 09-Jul-2004  
C:Accession: S05542  
R:Fuerst, P.; Moesch, H.U.; Solioz, M.  
Nucleic Acids Res. 17, 6724, 1989  
A:Title: A protein of unusual composition from Enterococcus faecium.  
A:Reference number: S05542; PMID:89385998; PMID:2780297  
A:Accession: S05542  
A:Molecule type: DNA  
A:Residues: 1-507 <FUE>  
A:Cross-references: UNIPROT:PI3692; UNIPARC:UPI000016f6FC; GB:X16421; EMBL:M26048; NID:9  
A:Note: the authors translated the codon CGT for residues 221 and 223 as Lys  
Query Match 25.4%; Score 537.5; DB 2; Length 507;  
Best Local Similarity 28.7%; Pred. No. 8.7e-20;  
Matches 146; Conservative 89; Mismatches 172; Indels 101; Gaps 12;  
QY 10 LVSGVTLSSATLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTQGOVSALOTQ 69  
Db 1 MLSSIALTAVGSPIAAAADDFDSQIQQDKKIADLQNOQASQSOIEALEGGQVSAINTKA 60  
QY 70 ASLQENORLEAQSATLGGQIOTLSSKIVARNESLKQARSQAOKSNAATSYINAIINSKS 129  
Db 61 QDLLTKQDTLRKESALQKEIKDLQERIEKRAATQKQARETOVKNTSYNIDAVLNADS 120  
QY 130 VSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENQAAINTVAANOQETIAQNTNAL 189  
Db 121 LADAVGRIGAMSTIVKANQDLVQQKEDQKQAEAKKAENEAKQKELADNQAALQESQKGL 180  
QY 190 NTQQAQLEAAQNLQAEELTTAODQKATLVAQKA-----AAEEAARQAAQAQ 235  
Db 181 LAKQADNLVLTSLAAEQATSDKADLNRKKAEEAEQARIREQARLAEQARQQAQAEK 240  
QY 236 AAAEAKAAAEAKALQEQAAQAQAAANNNNTQATDASDQQA-----AADNTQ 282  
Db 241 AEKAREQAQAQAQATQALSSASTTTESSSAQSSEESKAPESSTTEBESTESTSTTTE 300  
QY 283 AAQTGDSTDQSA-----AQAVNNSDQESTTATAAQPSSASSASTAAVAANTSSANT----- 332

Db 301 NSSTGSSSTESSTBESTVPESTQESTPANTPESSSSSNTNNTNNTNNTNNTNNT 360  
QY 333 -----YPAQOCTMGVKSLA-----PWMGNTWNG----- 356  
Db 361 NNNNNNTVTPTPTPTPAPAPAPNPSPSVNGAIVAEAYKYGTPYV---WGGKDPG 417  
QY 357 -----GOWAASAAAAGYRVG-STPSAGAVAVVND-GGYGHVAYVTG 395  
Db 418 FDCSGTRVYVLQVTRDIGMTVPQESAGTKISVSQAKAGDLLFWGSAGTYHVAISLG 477  
QY 396 VGGQIQIQVEANYAGNQSIGNVGFNFP 423  
Db 478 --GGQ-YIHAPQGENVKGVQ-WYTP 501  
RESULT 6  
AC1763  
peptidoglycan lytic protein P45 [imported] - Listeria innocua (strain Clp11262)  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AC1763  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Matok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AC1763  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-398 <GLA>  
A:Cross-references: UNIPROT:Q927Y8; UNIPARC:UPI00000CC935; GB:AL592022; PIDN:CAC97875.1;  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: spl  
Query Match 16.8%; Score 356; DB 2; Length 398;  
Best Local Similarity 27.1%; Pred. No. 6.2e-11;  
Matches 114; Conservative 87; Mismatches 147; Indels 72; Gaps 16;  
QY 1 MKKRILSAV-LVSGVTLSSATLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTQ 59  
Db 1 MKKNTFIAISLAADVSLTPAPTNNVFA--DVNTDIQNDQKINDIKSKKTGLQSLSSLV 58  
QY 60 GOVSALQTOQAELQAEORLEAQSATLGGQIOTLSSKIVARNESLKQARSQAOKSNAATS 119  
Db 59 ADLEKAQEKAKSLQEGFQDTGKELQNLQDIDINERIKERETVLKERARAWOKTSNSNA 118  
QY 120 YINAIINSKVSDAINRVSAIREVVSANEKMLHQEQD-----KAAVEQKHQENQAAI 172  
Db 119 YLEVLIDRAENLSDLVGRVSAVNVQLVDSKLSILEDQKDEKALKTKQTAVKKQEQAATAI 178  
QY 173 NTVAANOQETIAQNTNALNTQQAQLEAAQNLNLOAELTTAODQKATLVAQKAAAEAEARQAA 232  
Db 179 HEFEAQQ-----NKIEAQKAEKAIVAQLAADQASAEENEKAGLVSE--DKAAKE-- 226  
QY 233 AAQAAAEAKAAAEAKALQEQAAQAQAAANNNNTQATDASDQQAADANTQAAQTGDSTDQ 292  
Db 227 -----ATARATALRE-----ATDANVGQOTT--NTNA--SSNSKTS 258  
QY 293 SAAQAVNNSDQESTTATAAQPSSASSASTAAVAANTSSANTYPAG-----OCTWGVKSL 345  
Db 259 NKVESTNNSEAPS-----HATPSGGGYSAMIAARAQLGKPYSLGATGPSAFDCS-GFTSY 313  
QY 346 ---APWVGNWNGWGOWAASAAAAGYRVGSTPSAGAVAVN-DGGYGHVAYVTGQGGQI 401  
Db 314 APRAAGVSLPRTSGGQYAAASKIS-----ASQAKPGDLVFFNYGGGTAHVGIYVG--GGQM 367  
RESULT 7

Al1387  
peptidoglycan lytic protein p45 [imported] - Listeria monocytogenes (strain EGD-e)  
C:Species: Listeria monocytogenes  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: Al1387  
R:Glaser, P.; Prangeul, L.; Buchrieser, C.; Anand, A.; Baquero, F.; Berche, P.; Bloeker  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
J.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Krett, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A.; Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: Al1387  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-401 <GLA>  
A:Cross-references: UNIPROT:Q9RE04; UNIPARC:UPI00000D019D; GB:NC\_003210; PIDN:CAD00583.1  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: spi

Query Match 16.2%; Score 343.5; DB 2; Length 401;  
Best Local Similarity 26.4%; Pred. No. 2.6e-10;  
Matches 111; Conservative 92; Mismatches 148; Indels 69; Gaps 15;  
QY 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTIQ 59  
DB 1 MKKNTFIAISLAAVISLTPAFTTNVFA--DVNTDIQNQDKINDIKSKKTDLQSLGLIV 58  
QY 60 GQVSALQTOQAELQAEORLEAQSATLGOIOTLSSKTVARNESLKQOARSQAOKSNAATS 119  
DB 59 ADLEAQAQKAKSLQGEFKTKGKELKLNEDIKSIINERIKERTVLKERARAKQTSNSNA 118  
QY 120 YINAIINSKVSDAINRVSASREVSANERKMLHQEQDKAAVE-----QKHQENQA-AI 172  
DB 119 YLEVILDAENLSDLGRVSAVNQLVSDKSILEDQONDEKALKTKQTAVKKKQEQDQATAI 178  
QY 173 NTVAAQETIAQNTALNTQQAQLAAQLNLAELTTAQDQKATVAQKAAEAEARQAA 232  
DB 179 HEYEAQQ-----NKIEAQAQKAEKAIVAQLASDQAENAKAGLVSR---DKAAKE-- 226  
QY 233 AAQAAAEAKAAEAKALQEOAAQAAQAAANNNTQATDASDQAAAAAATQAAQTGDSTDQ 292  
DB 227 -----ATARATALRE-----ATSSNVGDESSSDTSPSKSNITK 262  
QY 293 SAAQAQVNSDQBSTTTATAAQAQSSASTAAVAANTSSANTYPAG-----QCTWGVKSL 345  
DB 263 NVASNDNS-----APSAATPSSGGYSAMIISAANAQLGPKYSLGATGPSAFDCS-GFTSY 316  
QY 346 ---APWVNGYWGNGQWAAASAAAAGRYRGSTPSAGAVAVNDG-CYGHVAVYTVGGQOI 401  
DB 317 AFRAAGVSLPRTSGQGYAAASKIS----ASQAKPGDLVFFNNGSGIAHVGIVYG--GGQM 370

RESULT 8  
F70031  
cell wall-binding protein homolog yvcE - Bacillus subtilis  
C:Species: Bacillus subtilis  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C:Accession: F70031  
R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
C.; Bron, S.; Brouillet, S.; Brucher, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.F.; Zumstein, B.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: F70031  
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-473 <KUN>  
A:Cross-references: UNIPROT:P40767; UNIPARC:UPI0000060AE1; GB:Z99121; GB:AL009126; NID:G  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yvcE

Query Match 15.1%; Score 319; DB 2; Length 473;  
Best Local Similarity 23.9%; Pred. No. 5.1e-09;  
Matches 111; Conservative 94; Mismatches 187; Indels 72; Gaps 11;  
QY 1 MKKRILSAVLVSGVTLS-----ATLTSVAVKADDFDAQIASQDSKINNLTAAQOAAQAAQ 54  
DB 1 MKKSLITLGLASVIGTSSFLIPFTSKTASAEFLDEKKQKIESQSEV---ASSIEAKEKE 57  
QY 55 VNTQGVSAQTQQAEL-----QAENQRLAQSATLGOIOTLSSKIVAR 100  
DB 58 LTELQENQSKIEKELKINDKALDTSNKIEDKEENDTKKEIKLKEIKETEAREIEKR 117  
QY 101 NESLKQOARSQAOKSNAATSYINAIINSKVSDAINRVSASREVSANERKMLHQEQDKAA 160  
DB 118 NEILKRVRSIQESGGSGYIDVLGTSFGDFISRTAVSSIVDADKDLIKQQEQDKAK 177  
QY 161 VE-----QKHQENQAANTVAANQETIAQNTALNTQQAQLAAQLNLAELTTAQDQ 213  
DB 178 LEDSEADLNDKLKEVQAALAKLETWQKDLQNLNEKDKLFDEAKASQKTKAKAISLKESE 237  
QY 214 KATLVAQKAAEAEARQAAQAAEAKAAEAKALQEOAAQAAQAAANNNTQATDASDQ 273  
DB 238 ASELANQKANTE-----AEQARIKKEQEAALAIKKQBEAQKA-----SDETQTDSD 284  
QY 274 QAAADNTQAAQTGDSTDQSAQAQVNSDQBSTTTATAAQAQSSASTAAVAANTSSANTY 333  
DB 285 QTATTESSKASDSDSDNSDSSNGSSNGSSNGSSNGSSNGSSNGSGTVISNSGGIE 344  
QY 334 PAGQCTWGVKSLAPWVNGYWGNG-----QWA-ASAAAAGRYRGSTPS 375  
DB 345 GAISVSGSIVGQSP----YKFGGRTQSDINNRIFDSCSFVRWAYASAGVNLGPVGTTT 400  
QY 376 AGAVAVNDGGYGHVAVYTVGGQIOVQEAANYAGNQSIGNYRG 419  
DB 401 DTLV-----GRGQAVSASEMKRGDL-VFFDPTKTNHGVGYLG 437

RESULT 9  
E75383  
conserved hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: E75383  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; i  
S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zaleski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.W.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: E75383  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-581 <WHI>  
A:Cross-references: UNIPROT:Q9RU45; UNIPARC:UPI00000C195A; GB:AE001998; GB:AE000513; NID:  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR1549  
A:Map position: 1  
Query Match 12.3%; Score 259.5; DB 2; Length 581;



QY 347 PWGNYWGN-----GGWAASAAAGYRVCGSTPSAGAVVWNDGGYCHVAYVTGVGG 399

Db 323 PVTGKYESHKGQDIAGGCTVTVSAAASGTVFSGFGASGSGF---GGYGVVVKIDHNGGF 379

QY 400 Q 400

Db 380 Q 380

RESULT 12

AB1763

cell wall binding protein homolog lin2647 [imported] - Listeria innocua (strain Clip1126

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004

C:Accession: AB1763

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreitt, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AB1763

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-437 <GLA>

A:Cross-references: UNIPROT:Q927Y9; UNIPARC:UPI00000CC934; GB:AL592022; PIDN:CAC97874.1;

A:Experimental source: strain Clip11262

C:Genetics:

A:Gene: lin2647

Query Match 11.9%; Score 251; DB 2; Length 437;

Best Local Similarity 23.2%; Pred. No. 1.1e-05;

Matches 101; Conservative 80; Mismatches 149; Indels 106; Gaps 16;

QY 6 LSAVLVSGVTLSATTLSAVKAD-----DFAQIASQDSKINNLTAAQ 47

Db 11 LSLIIIS-----APLTSYKAESINDMKQROSEIEQKKSELNKLDTKNSLHNLENAE 63

QY 48 QAAA---QAOVNTIQGVSAQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESL 104

Db 64 KDAKELESLSIDETNKKLKEQDKVDSENEKL-----KLKKEIEKLNDIRQKVL 119

QY 105 KQARSQAQKNAATSYINAININSKVSADINRVSAIREVVSANKEKVLHQEQDK---AAV 161

Db 120 DSRARAIQTGTATSYDLMIPEADDFKELIDRTVTVVSAIVKADQNMQDKDDQKCLKVA 179

QY 162 FQKHQENQAAINTVAANOETTAQNTNALNTQQAQLEAAQLNLAELTTAQDQKATLVQAK 221

Db 180 ENSSEKKLENLKVLAVELEVSNNMESQKQKNDLVMALAN-KDOLT--KSEQTLLTNEQ 236

QY 222 AAABEAAQAAAAAQAABAAKAAAEAKALQEAQAAQAAANNNNTQATDASDQAAAAADNT 281

Db 237 GALSDEQKLLANIGAKQKAEALKAABEKMQAAXN----- 276

QY 282 QAAQTGSDTQSAQAQVNNSDQESTTATAQPSA--SSASTAAVAANTSSANTY--PA-G 336

Db 277 -----ATVAAQQPSVTSAGTGATDTVSSGGQGFIPKPSG 312

QY 337 QCTWGVKSLA-PWGNVWGN-----CGOWAASAAAAGYRV-----GSTPSAGAVVWVD 384

Db 313 MLTSGFSERTNPVTGKYESHKGQDIAGGCTITVSAASGRVVFSFGATGS-----GF 365

QY 385 GGYGHVAVVTGVGGQ 400

Db 366 GGYGVVVKIDHNGFGQ 381

RESULT 13

B75310

conserved hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C:Accession: B75310

R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; F

, M.; Shen, H.O.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, M.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: B75310

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-528 <WHI>

A:Cross-references: UNIPROT:Q9RSJ1; UNIPARC:UPI00000C1A76; GB:AE000513; NID

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR2133

A:Map position: 1

Query Match 11.2%; Score 237; DB 2; Length 528;

Best Local Similarity 29.0%; Pred. No. 6.4e-05;

Matches 94; Conservative 54; Mismatches 126; Indels 50; Gaps 14;

QY 5 ILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAAQAVNTIQGVSA 64

Db 48 VLSGMAISAASLAAPLLL-----NRNAVSTIAAQADQLRPQIEALRTEVGTVOGELRA 99

QY 65 LQTOQAELQAEORLEAQSATLGGQIQTLSSKIVARNESL-----KQARSQAQKNAATSY 120

Db 100 ARTER-----EAARSEAQKA--GQREAAARQELAAARQNLASAQEQEARLTKQAQDLQTR 152

QY 121 INAININSKVSADINRVSAIREVVSANKEKVLHQEQDKAAVEQKHQENQAAINTVAANO 180

Db 153 LKTUAEQRRLQEA--QAQAREKQLQASQKQL--QASEDRATQLDSQVLDLKRSQAQEQ 208

QY 181 TIAQNTNA-LNTTQQAQLEAAQLNLAELTTAQ--DQKATLVQAK-----AAAAEAAARQAA 232

Db 209 --AQNAQTFRANAQAARTEELQRRRAAAQATAQAQATRAAQAQASQAQASARAQEVREQAR 266

QY 233 AAQAAAEKAAAEAKALQEAQAAQAAANNNNTQATDASDQAAAAADNTQAAATGSDTDQ 292

Db 267 QAQRRAE---QAQARAQEVQ-AQAQAAA-----QASVROQAQAAQATQLGQVRTGAAQ 314

QY 293 SAAQAVNNSDQESTTATAQPSAS 316

Db 315 QVAAQ-----QQAQAAHRRPSSA 333

RESULT 14

JV0057

tolA protein - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 09-Jul-2004

C:Accession: JV0057; B64810

R:Levengood, S.K.; Webster, R.E.

J. Bacteriol 171, 6600-6609, 1989

A:Title: Nucleotide sequences of the tolA and tolB genes and localization of their produc

A:Reference number: JV0057; MUID:90078104; PMID:2687247

A:Accession: JV0057

A:Molecule type: DNA

A:Residues: 1-421 <LEV>

A:Cross-references: UNIPROT:P19934; UNIPARC:UPI0000137105; GB:M28232; NID:g148018; PIDN:f

A:Experimental source: strain JM105

A:Note: The authors translated the initiation codon GTG for residue 1 as Val

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

.A.; Rose, D.J.; Mau, B.; Shaao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: B64810

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-421 <BLAT>

A;Cross-references: UNIPARC:UPI0000137105; GB:AE000177; GB:U00096; NID:gl786955; PIDN:AA  
A;Experimental source: strain K-12, substrain MG1655  
C;Comment: tolA and tolB proteins are necessary for colicins E2, E3, A, and K to reach b  
C;Genetics:  
A;Gene: tolA  
A;Map position: 17 min  
A;Start codon: GTG  
C;Keywords: nucleotide binding; P-loop; transmembrane protein  
F;14-34/Domain: transmembrane #status predicted <MSS>  
F;78-301/Domain: helical #status predicted <HSR>  
F;355-362/Region: nucleotide-binding motif A (P-loop)

Query Match 10.9%; Score 230; DB 2; Length 421;  
Best Local Similarity 28.5%; Pred. No. 0.00011;  
Matches 117; Conservative 65; Mismatches 179; Indels 50; Gaps 19;

QY 5 ILSAVLVSGVTLSSATTLISAVKADDFDAQIASQ-----DSKINNLTAQQAQAQVNTIQ 59  
DB 16 IISAVL--HVILFAALIWS-----FDENIEASAGGGGSSIDA VVWDSGAVVEQYKRMQ 68  
QY 60 GQVSA-----LQTQQA--ELQ-----AENQL--EAQSATLGGQIQTLSSKIVARNES 103  
DB 69 SQESSAKSDEQKMKQQAABELEKQAAEQERLQLEKERLAAQEQKQAEAAKQAE 128  
QY 104 LKQ-QARSAQKSNATSYINAIINSKVSDDAIVNSAIREVVVANEKMLHQEQDKAAVE 162  
DB 129 LKQQAEEAAAKAAADAKAEADAKAEAAEAAK-----AAADAKKAEAAEAAKAAAE 181  
QY 163 -QKHQENQAAI--NTVAANOETIAQNTNALNTQQAQLEAAQLNLAEL--TTAQQDOKATIV 218  
DB 182 AQKKAEEAAALUKKAEAAEAAAEAAAEARKKAATEAAEKAKAEKAAAEKAAADKKA--A 239  
QY 219 AQKAAAE--EAAQAAAAQAQAAEAKAAAEAKALQEQAQAQAAANNNNT--QATDASDQA 275  
DB 240 AEKAAADKKAEEAAEAAEAAADKKAEEKAAAEKAAAEKAAAEKAAAEAAEADDIFG 299  
QY 276 AAADNTQAAQTGSDTQGAQAQVNNSDQESTTATAQ-PSASSASTAAVAANTSSANTYP 334  
DB 300 ELSGKNAPKTGGGAKGNASPAAGSNTKNGGASGADINNYAGQIKSAIESKFYDASSYA 359  
QY 335 AGCQTMGVKSLAP--WVGNWNGGOWA--ASAAAGYRVGTPSAGAVAVW 382  
DB 360 GKTCTLRIRK-LAPDGMILLDIKPEGGDPAALCQAALAAAKLAKIPKPPSQAVY 409

RESULT 15  
B90835  
probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, substrain RIM  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C;Accession: B90835  
R;Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
A;Reference number: A39629; PMID:21156231; PMID:11258796  
A;Accession: B90835  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-971 <HAY>  
A;Cross-references: UNIPROT:Q8XDQ4; UNIPARC:UPI000016542B; GB:BA000007; PIDN:BA035073.1;  
A;Experimental source: strain O157:H7, substrain RIMD 0509952  
C;Genetics:  
A;Gene: EC81650

Query Match 10.8%; Score 228; DB 2; Length 971;  
Best Local Similarity 26.4%; Pred. No. 0.00034;  
Matches 96; Conservative 65; Mismatches 159; Indels 44; Gaps 10;

QY 19 ATTLISAVKADDFDAQIASQDSKINNLTAQQAQAQVNTIQGVSAALQTQQAELQAEQNR 78  
DB 121 AQNTAAKKASDASTSAREATHATDRAADSARAASTSGAQAASSAQSSAGTASTKA 180

Search completed: February 15, 2006, 18:10:24  
Job time : 22.0162 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.  
OM protein - protein search, using sw model  
Run on: February 15, 2006, 17:57:21 ; Search time 123.762 Seconds  
(without alignments)  
2462.693 Million cell updates/sec  
Title: US-10-797-821-32  
Perfect score: 2116  
Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNVRGWNPGSVIYIPN 432

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 2166443 seqs, 705528306 residues  
Total number of hits satisfying chosen parameters: 2166443  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : UniProt\_05.80.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match % | Length | DB | ID            | Description        |
|------------|--------|---------------|--------|----|---------------|--------------------|
| 1          | 2116   | 100.0         | 432    | 2  | Q938V1_STRMU  | Q938v1 streptococc |
| 2          | 2099   | 99.2          | 432    | 2  | Q938V2_STRMU  | Q938v2 streptococc |
| 3          | 2074.5 | 98.0          | 431    | 2  | Q9AG98_STRMU  | Q9ag98 streptococc |
| 4          | 2070.5 | 97.8          | 431    | 2  | Q938V3_STRMU  | Q938v3 streptococc |
| 5          | 2070.5 | 97.8          | 431    | 2  | Q8DWM3_STRMU  | Q8dwm3 streptococc |
| 6          | 2066.5 | 97.7          | 431    | 2  | Q938V0_STRMU  | Q938v0 streptococc |
| 7          | 1229.5 | 58.1          | 447    | 2  | Q9AKA4_STRAG  | Q9aka4 streptococc |
| 8          | 1229.5 | 58.1          | 447    | 2  | Q8E2H1_STRAS  | Q8e2h1 streptococc |
| 9          | 1229.5 | 58.1          | 447    | 2  | Q8E7X9_STRAS  | Q8e7x9 streptococc |
| 10         | 1087.5 | 51.4          | 474    | 2  | Q5M6K4_STRTP2 | Q5m6k4 streptococc |
| 11         | 1087   | 51.4          | 398    | 2  | Q5XEL1_STRTP6 | Q5xel1 streptococc |
| 12         | 1087   | 51.4          | 398    | 2  | Q7CNQ7_STRP8  | Q7cnq7 streptococc |
| 13         | 1087   | 51.4          | 398    | 2  | Q8F318_STRP3  | Q8f318 streptococc |
| 14         | 1087   | 51.4          | 485    | 2  | Q5M212_STRT1  | Q5m212 streptococc |
| 15         | 1081   | 51.1          | 398    | 2  | Q9A1Z8_STRPY  | Q9a1z8 streptococc |
| 16         | 1053   | 49.8          | 392    | 2  | Q8DMY4_STRR6  | Q8dmv4 streptococc |
| 17         | 1053   | 49.8          | 392    | 2  | Q97N55_STRPN  | Q97n55 streptococc |
| 18         | 937    | 44.3          | 211    | 2  | Q9ZAS7_STRMU  | Q9zas7 streptococc |
| 19         | 674.5  | 31.9          | 461    | 1  | US445_IACLC   | P22865 lactococcu  |
| 20         | 658    | 31.1          | 456    | 2  | Q9CDJ1_LACLA  | Q9cdj1 lactococcu  |
| 21         | 585.5  | 27.7          | 524    | 2  | Q9K2J9_ENTFC  | Q9k2j9 enterococcu |
| 22         | 569.5  | 26.9          | 516    | 1  | P54_ENTFC     | P13692 enterococcu |
| 23         | 549    | 25.9          | 482    | 2  | Q5M5M6_STRTP2 | Q5m5m6 streptococc |
| 24         | 543.5  | 25.7          | 576    | 2  | Q9KJJ3_ENTHR  | Q9kjj3 enterococcu |
| 25         | 542    | 25.6          | 470    | 2  | Q8RQE3_GLACT  | Q8rqe3 lactococcu  |
| 26         | 530    | 25.0          | 482    | 2  | Q93LK4_ENTFA  | Q93lk4 enterococcu |
| 27         | 507.5  | 24.0          | 461    | 2  | Q5GSA7_STRTR  | Q5gsa7 streptococc |
| 28         | 498.5  | 23.6          | 449    | 2  | Q93LK3_ENTFA  | Q93lk3 enterococcu |
| 29         | 436.5  | 20.6          | 211    | 2  | Q8DVU8_STRMU  | Q8drv8 streptococc |
| 30         | 416    | 19.7          | 226    | 2  | Q93RG6_STRIT  | Q93rg6 streptococc |
| 31         | 415    | 19.6          | 544    | 2  | Q840X3_STRMU  | Q840x3 streptococc |

RESULT 1  
Q938V1\_STRMU Q938v1 STRMU PRELIMINARY; PRT; 432 AA.  
AC Q938V1.  
DT 01-DEC-2001 (TremBLrel. 19, Created)  
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)  
DE 01-MAR-2004 (TremBLrel. 26, Last annotation update)  
DE Glucan-binding protein B.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus  
OX NCBI\_TaxID=1309;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=3SNI;  
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;  
RT "Cloning of the gbpB gene from Streptococcus mutans.";  
RL J. Dent. Res. 79:224-224 (2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=3SNI;  
RX MEDLINE=21481971; PubMed=11598068;  
RX DOI=10.1128/IAI.69.11.6931-6941.2001;  
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,  
RA Duncan M.J.;  
RT "Cloning of the Streptococcus mutans gene encoding glucan binding  
RT protein B and analysis of genetic diversity and protein production in  
RT clinical isolates.";  
RL Infect. Immun. 69:6931-6941 (2001).  
DR EMBL; AY046413; AAK94503.1; -; Genomic\_DNA.  
DR InterPro; IPR007921; CHAP.  
DR InterPro; IPR009148; SIBA.  
DR Pfam; PF05257; CHAP; 1.  
DR PRINTS; PR01852; SIBAPROTEIN.  
DR PROSITE; PS50911; CHAP; 1.  
SQ SEQUENCE 432 AA; 44652 MW; 3F88ECB9A1F3BE4F CRC64;  
Query Match 100.0%; Score 2116; DB 2; Length 432;  
Best Local Similarity 100.0%; Pred. No. 1e-88;  
Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKKRILSAVLVSGVTLSATTL SAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60  
Db 1 MKKRILSAVLVSGVTLSATTL SAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60  
QY 61 QVSALQTOQAELOAENRLEAQSATLGGQIQTLSSKIVARNESLKQOARSACKSNAATSY 120  
Db 61 QVSALQTOQAELOAENRLEAQSATLGGQIQTLSSKIVARNESLKQOARSACKSNAATSY 120  
QY 121 INAININSKVSDAINRVSAIREVVSANEKMLHQEQDKAAVEQKHQENQAATNTVAANQE 180  
Db 121 INAININSKVSDAINRVSAIREVVSANEKMLHQEQDKAAVEQKHQENQAATNTVAANQE 180  
QY 181 TTAQNTNALNTQQAOLEAAQLNLQALTTAQQKATLVAQKAAAEAEARQAQAAAEAA 240

```

Db      181  TTAQNTNALNTTQAQLEAAQLNLQAELETTTAQDQKATLVAQKAAAEARQAQAAAEAA 240
QY      241  KAAAEAKALQEQAAQAAQAAANNNTTQATDASDQQAADNTQAAQTGSDTDSAAQAVNN 300
Db      241  KAAAEAKALQEQAAQAAQAAANNNTTQATDASDQQAADNTQAAQTGSDTDSAAQAVNN 300
QY      301  SQESTTTATAOPSSASSASTAAVAANTSSANTYPAGQCTGWKSLAPWVGNWGGQWA 360
Db      301  SQESTTTATAOPSSASSASTAAVAANTSSANTYPAGQCTGWKSLAPWVGNWGGQWA 360
QY      361  ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
Db      361  ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
QY      421  FNPGSVSYIYPN 432
Db      421  FNPGSVSYIYPN 432

RESULT 2
Q938V2 STRMU
ID Q938V2 STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V2;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=15JP2;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=15JP2;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL; AY046412; AAK94502.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44648 MW; E769B2504AE50E9 CRC64;

Query Match 99.2%; Score 2099; DB 2; Length 432;
Best Local Similarity 99.3%; Pred. No. 6.1e-88;
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1  MKKRLTSLAVLNSGVLTSSNTTILSAVKADDFDAQIASQDSKINNLTAQQAQAAQVNTIQ 60
Db      1  MKKRLTSLAVLNSGVLTSSNTTILSAKADDFDAQIASQDSKINNLTAQQAQAAQVNTIQ 60
QY      61  QVSALQTOQAEIQAENQRLQEAQSATILGQOIQTLSKIVARNESLQQAARSQAQNAATSY 120
Db      61  QVSALQTOQAEIQAENQRLQEAQSATILGQOIQTLSKIVARNESLQQAARSQAQNAATSY 120
QY      121  INAINSKSVSDAINRVSAIREVWSANEKMLHQEQDQKAAVEQKHQENQAANTVAANQE 180
Db      121  INAINSKSVSDAINRVSAIREVWSANEKMLQQEQDQKAAVEQKHQENQAANTVAANQE 180

```

## RESULT 3

```

Q9AG98 STRMU
ID Q9AG98 STRMU PRELIMINARY; PRT; 431 AA.
AC Q9AG98;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-FEB-2005 (TremBLrel. 29, Last annotation update)
DE Immunodominant glycoprotein IDG-60 (Glucan-binding protein B).
GN Name=saga;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=21153617; PubMed=11254612;
RX DOI=10.1128/IAI.69.4.2493-2501.2001;
RA Chia J.S., Lee Y.Y., Huang P.T., Chen J.Y.;
RT "Identification of stress-responsive genes in Streptococcus mutans by
RT differential display reverse transcription-PCR.";
RL Infect. Immun. 69:2493-2501(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=21481977; PubMed=11598074;
RX DOI=10.1128/IAI.69.11.6987-6998.2001;
RA Chia J.S., Chang L.Y., Shun C.T., Chang Y.Y., Chen J.Y.;
RT "A 60-kilodalton immunodominant glycoprotein is essential for cell
RT wall integrity and the maintenance of cell shape in Streptococcus
RT mutans.";
RL Infect. Immun. 69:6987-6998(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RA Chia J.-S., Chang L.-Y., Lee Y.-Y., Chen J.-Y.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3VF4;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3VF4;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in

```



```

RT clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL: AF338445; AAK08104.1; -; Genomic DNA.
DR EMBL: AY046411; AAK94501.1; -; Genomic_DNA.
DR InterPro: IPR007921; CHAP.
DR InterPro: IPR009148; SIDA.
DR Pfam: PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
SQ SEQUENCE 431 AA; 44592 MW; 3EB221FC5E47232E CRC64;

Query Match 98.0%; Score 2074.5; DB 2; Length 431;
Best Local Similarity 98.8%; Pred. No. 7.9e-87;
Matches 427; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
DB 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
QY 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
DB 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
QY 121 INAIINSKSVSDAINRVSAIREVVSANEKMLHQOEQDKAAVEQKHQENQAANTVAANO 180
DB 121 INAIINSKSVSDAINRVSAIREVVSANEKMLHQOEQDKAAVEQKHQENQAANTVAANO 180
QY 181 TTAQNTNALNTQQAOLEAAQLNLQAEELTTAQQDKATLVAKAAAEAAEAQAQAAQAAEA 240
DB 181 TTAQNTNALNTQQAOLEAAQLNLQAEELTTAQQDKATLVAKAAAEAAEAQAQAAQAAEA 240
QY 241 KAAAEAKALQEQAAQAQAAANNNNTQATDASDQQAADNTQAAQTGSDTDSAAQAVNN 300
DB 241 KAAAEAKALQEQAAQAQAAANNNNTQATDASDQQAADNTQAAQTGSDTDSAAQAVNN 300
QY 301 SQESTTATAQPSASSASTAAVAANTSSANTYPAGCTGWKSLAPWVGNWNGGQWA 360
DB 301 SQESTTATAQPSASSASTAAVAANTSSANTYPAGCTGWKSLAPWVGNWNGGQWA 360
QY 361 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
DB 361 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
QY 421 FNPGSVSYIYPN 432
DB 421 FNPGSVSYIYPN 431

RESULT 4
Q938V3 STRMU PRELIMINARY; PRT; 431 AA.
AC Q938V3
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OC NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN-SJ32;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC MEDLINE=21481971; PubMed=11598068;
RX STRAIN-SJ32;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RX Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;

RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL: AY046410; AAK94500.1; -; Genomic_DNA.
DR InterPro: IPR007921; CHAP.
DR InterPro: IPR009148; SIDA.
DR Pfam: PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
SQ SEQUENCE 431 AA; 44620 MW; 464FE3B563FB7E51 CRC64;

Query Match 97.8%; Score 2070.5; DB 2; Length 431;
Best Local Similarity 98.6%; Pred. No. 1.2e-86;
Matches 426; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
DB 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
QY 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
DB 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
QY 121 INAIINSKSVSDAINRVSAIREVVSANEKMLHQOEQDKAAVEQKHQENQAANTVAANO 180
DB 121 INAIINSKSVSDAINRVSAIREVVSANEKMLHQOEQDKAAVEQKHQENQAANTVAANO 180
QY 181 TTAQNTNALNTQQAOLEAAQLNLQAEELTTAQQDKATLVAKAAAEAAEAQAQAAQAAEA 240
DB 181 TTAQNTNALNTQQAOLEAAQLNLQAEELTTAQQDKATLVAKAAAEAAEAQAQAAQAAEA 240
QY 241 KAAAEAKALQEQAAQAQAAANNNNTQATDASDQQAADNTQAAQTGSDTDSAAQAVNN 300
DB 241 KAAAEAKALQEQAAQAQAAANNNNTQATDASDQQAADNTQAAQTGSDTDSAAQAVNN 300
QY 301 SQESTTATAQPSASSASTAAVAANTSSANTYPAGCTGWKSLAPWVGNWNGGQWA 360
DB 301 SQESTTATAQPSASSASTAAVAANTSSANTYPAGCTGWKSLAPWVGNWNGGQWA 360
QY 361 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
DB 361 ASAAAAGYRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
QY 421 FNPGSVSYIYPN 432
DB 421 FNPGSVSYIYPN 431

RESULT 5
Q8DWM3 STRMU PRELIMINARY; PRT; 431 AA.
AC Q8DWM3
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative secreted antigen GbpB/SagA; putative peptidoglycan
DE hydrolase.
GN Name=gbpB; OrderedLocusNames=SMU.22;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OC NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN-UA159 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";

```

```

RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).
DR EMBL; AE014855; RAN57811.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 431 AA; 44620 MW; 2D1CA685248CCD3E CRC64;

Query Match      97.8%; Score 2070.5; DB 2; Length 431;
Best Local Similarity 98.6%; Pred. No. 1.2e-86;
Matches 426; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGOOIOTLSSKIVARNESLKQOARSQKSAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGOOIOTLSSKIVARNESLKQOARSQKSAATSY 120
QY 121 INAIINSKSVSDAINRVSAIREVVSANERKMLHQEQDKAAVEQKHQENQAINTVAAHQE 180
Db 121 INAIINSKSVSDAINRVSAIREVVSANERKMLHQEQDKAAVEQKHQENQAINTVAAHQE 180
QY 181 TIAQNTNALNTQQAOLEAAQLNLQAEELTTAQQOKATLVAQKAAABEAAARQAAAAQA 240
Db 181 TIAQNTNALNTQQAOLEAAQLNLQAEELTTAQQOKATLVAQKAAABEAAARQAAAAQA 240
QY 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAOTGSDTQSAQAQVNN 300
Db 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAOTGSDTQSAQAQVNN 300
QY 301 SDQESTTATAAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGQWA 360
Db 301 SDQESTTATAAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGQWA 360
QY 361 ASAAAAGRVGSTPSAGAVAVVNDGGYGHVAVVTGQGGQIQVQENYAGNQSIGNYRGW 420
Db 361 ASAAAAGRVGSTPSAGAVAVVNDGGYGHVAVVTGQGGQIQVQENYAGNQSIGNYRGW 420
QY 421 FNPGSVSYIYPN 432
Db 421 FNPGSVSYIYPN 431

RESULT 6
Q938V0_STRMU PRELIMINARY; PRT; 431 AA.
AC Q938V0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5SM3;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
[2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5SM3;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/JAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding

RT protein B and analysis of genetic diversity and protein production in
clinical isolates.";
RL EMBL; AY046414; AAK94504.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
SQ SEQUENCE 431 AA; 44650 MW; 05D38D8DB8C4609F CRC64;

Query Match      97.7%; Score 2066.5; DB 2; Length 431;
Best Local Similarity 98.4%; Pred. No. 1.8e-86;
Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGOOIOTLSSKIVARNESLKQOARSQKSAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGOOIOTLSSKIVARNESLKQOARSQKSAATSY 120
QY 121 INAIINSKSVSDAINRVSAIREVVSANERKMLHQEQDKAAVEQKHQENQAINTVAAHQE 180
Db 121 INAIINSKSVSDAINRVSAIREVVSANERKMLHQEQDKAAVEQKHQENQAINTVAAHQE 180
QY 181 TIAQNTNALNTQQAOLEAAQLNLQAEELTTAQQOKATLVAQKAAABEAAARQAAAAQA 240
Db 181 TIAQNTNALNTQQAOLEAAQLNLQAEELTTAQQOKATLVAQKAAABEAAARQAAAAQA 240
QY 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAOTGSDTQSAQAQVNN 300
Db 241 KAAAEKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAOTGSDTQSAQAQVNN 300
QY 301 SDQESTTATAAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGQWA 360
Db 301 SDQESTTATAAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGQWA 360
QY 361 ASAAAAGRVGSTPSAGAVAVVNDGGYGHVAVVTGQGGQIQVQENYAGNQSIGNYRGW 420
Db 361 ASAAAAGRVGSTPSAGAVAVVNDGGYGHVAVVTGQGGQIQVQENYAGNQSIGNYRGW 420
QY 421 FNPGSVSYIYPN 432
Db 421 FNPGSVSYIYPN 431

RESULT 7
Q9AKA4_STRAG PRELIMINARY; PRT; 447 AA.
AC Q9AKA4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PcsB protein precursor.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=6313;
RX MEDLINE=21101799; PubMed=11157929;
RX DOI=10.1128/JB.183.4.1175-1183.2001;
RX Reinscheid D.J., Gottschalk B., Schubert A., Eikmanns B.J.,
RA Chhatwal G.S.;
RT "Identification and molecular analysis of PcsB, a protein required for
cell wall separation of group B streptococcus.";
RL J. Bacteriol. 183:1175-1183 (2001).
DR EMBL; AJ277292; CAC28144.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.

```



```

DR Sagalist; qbs0016; -.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS09111; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 58.1%; Score 1229.5; DB 2; Length 447;
Best Local Similarity 59.7%; Pred. No. 2e-48;
Matches 276; Conservative 42; Mismatches 99; Indels 45; Gaps 9;

QY 1 MKKRILSAVLVSGVTLSSTATLSAVKADDDFAQIASQDSKINNLTAAQOAAQAVNTIQG 60
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 58
QY 61 QVSALQTOQAELEAENORLEAQSATLGGQIQITLSSKIVARNESLKQARSQAQSNATSY 120
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 117
QY 121 INAIINSKVSDAINRVSAIREVSNEMKMLHQBDKAAVEQKHQENQAANTVAANQE 180
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 177
QY 118 INTILNSKVSDAVNRVVAIREVSNEMKMLAQEADKAALEAKQIENQNAINTVVAANKQ 177
QY 181 TIAQNTNALNTQOALEAQLNQLAELTTAODKATLVAQKAAAEAAQAAQAAEA 240
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 237
QY 178 ALENKAALATQRAQLEAAQLELSAQITTVQKASLIQAKQAEAAKAAEAQAAEA 237
QY 241 KAAAEAKALQEQAAQAAANNNTQATDASDQQAADNTQAAQTGSDTQSAQAQAVNN 300
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 285
QY 238 KAQAEAKAQAEVAKAQAA-----QVESATPTETVQTPTEIKPSNLAT 285
QY 301 SDQ-----ESTTATA-----AQPSASASTAA-----VAANTSANTYPAG 336
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 345
QY 286 SSATTVAITTTATNEPKVTPQSVTVKAVEAPKVVSTPRAVSKPVVRSYSSNTYPMG 345
QY 337 OCTWGVKSLAPWGVNGWGQWASAAAGYRVGSTPSAGAVVW--NDGGYGHVAVYT 394
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 405
QY 346 OCTWGAKNASWGVNGWNGWASARAAGYSGVTTPRVGAVAVPVDYDGGYGHVAVYT 405
QY 395 GV-QGGQIQVQBANTYAGNQSIGNYRGWFPN---GSVSVIYPN 432
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 447
QY 406 SVANNSSIQVMSYAGNMSIGNYRGWFPNPSAGSVIYPN 447

RESULT 10
Q5M6K4_STRT2
ID Q5M6K4_STRT2 PRELIMINARY; PRT; 474 AA.
AC Q5M6K4
DT 01-FEB-2005 (TREMBLrel. 29, Created)
DT 01-FEB-2005 (TREMBLrel. 29, Last sequence update)
DT 13-SEP-2005 (TREMBLrel. 31, Last annotation update)
DE Glucan binding protein (Pc8B).
GN Name=pc8B; OrderedLocusNames=stcu022;
OS Streptococcus thermophilus (strain ATCC BAA-250 / LMG 18311).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=264199;
RN (1)
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15543133; DOI=10.1038/nbt1034;
RA Bolotin A., Quinquis B., Renault P., Sorokin A., Ehrlich S.D.,
RA Kulakauskas S., Lapidus A., Goldsman E., Mazur M., Pusch G.D.,
RA Fontein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D.,
RA Ngui K., Masuy D., Hancy F., Burtel S., Boutry M., Delcours J.,
RA Goffeau A., Hols P.;
RT "Complete sequence and comparative genome analysis of the dairy
RT bacterium Streptococcus thermophilus.";
RL Nat. Biotechnol. 22:1554-1558 (2004).
RN (2)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=LMG 18311;

RA Borges P., Layec S., Thibessard A., Fernandez A., Gintz B., Hols P.,
RA Decaris B., Leblond-Bourget N.;
RT "cse, a Chimeric and Variable Gene, Encodes an Extracellular Protein
RT Involved in Cellular Segregation in Streptococcus thermophilus.";
RL J. Bacteriol. 187:2737-2746 (2005).
DR EMBL; CP000023; AAV59752.1; -; Genomic DNA.
DR EMBL; AV730643; AAW82375.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS09111; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 474 AA; 48142 MW; 1BAC6A9E0A0A200D CRC64;

Query Match 51.4%; Score 1087.5; DB 2; Length 474;
Best Local Similarity 51.7%; Pred. No. 6.1e-42;
Matches 247; Conservative 72; Mismatches 108; Indels 51; Gaps 12;

QY 1 MKKRILSAVLVSGVTLSSTATLSAVKADDDFAQIASQDSKINNLTAAQOAAQAVNTIQG 60
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 57
QY 1 MKKRILSAVLVSGVTLSAA--ASVHAEDYDSQIAATNNAINSLASQOEAQAQVATIQS 57
QY 61 QVSALQTOQAELEAENORLEAQSATLGGQIQITLSSKIVARNESLKQARSQAQSNATSY 120
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 117
QY 58 QVSTLRTQKTELEAKNAELEKVSADLESEIQELSSKIVARQDSLAKQARSQAQNTATSY 117
QY 121 INAIINSKVSDAINRVSAIREVSNEMKMLHQBDKAAVEQKHQENQAANTVAANQE 180
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 177
QY 118 INSILNSKISSEIATITRITAIKSVTANNLLTKQESDQKELAAQOEAQNTTAAANKS 177
QY 181 TIAQNTNALNTQOALEAQLNQLAELTTAODKATLVAQKAAAEAAQAAQAA-----AQA 236
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 237
QY 178 ELETTEAGLTQAELEAAQVTLAELATQNEKTSLSAKSTAESVAASTAASVAQSOA 237
QY 237 AAEAKA-----AAEAKALQEQAAQAAANNNTQATDASDQQAADNTQAAQTG 287
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 296
QY 238 IAESATAQVWASSEATSVASSEVAATSEAVQSPVSETS-TASEAAQEPASSETS 296
QY 288 DSTDQSAQAQVNN-----SDQESTTATAQPS-----ASSASTAAV-----AANT- 327
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 356
QY 297 EVPESAAAPVSEAPASVAPVATSEAPATSEAPASVAPVATSEAPASVAPVATSEAPASV 356
QY 328 -----SSANTYPAGQCTWGVKSLAPWGVNGWGQWASAAAGYRVGSTPSAGAVVW 382
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 416
QY 357 KVSAASTENTTPVQCTWGVKSLAPWGVNGWGNKAKWIASAQAGHSVGTTPVAGIAVW 416
QY 383 -ND-GGYGHVAVYTGVOG-GQIQVQEAANYAGNQSIGNYRGWFPN-----GSVSVIYP 431
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 474
QY 417 PNDGGYGHVAVYTSAGANSIQVMSYAGNMSISNYRGTDPTTSSAHGSGSVFYIYP 474

RESULT 11
Q5XEL1_STRP6
ID Q5XEL1_STRP6 PRELIMINARY; PRT; 398 AA.
AC Q5XEL1
DT 25-OCT-2004 (TREMBLrel. 28, Created)
DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Secreted protein.
GN OrderedLocusNames=M6_Spy0017;
OS Streptococcus pyogenes (serotype M6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301450;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MGAS10394;
RX PubMed=15272401; DOI=10.1086/422697;
RA Banks D.J., Porcella S.F., Barbican K.D., Beres S.B., Phillips L.E.,
RA Voyich J.M., DeLeo F.R., Martin J.M., Somerville G.A., Musser J.M.;
RT "Progress toward characterization of the group A Streptococcus
RT metagenome: complete genome sequence of a macrolide-resistant serotype

```

```
RT M6 strain";
RL J. Infect. Dis. 190:727-738 (2004).
DR EMBL; CP000003; AAT86152.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match 51.4%; Score 1087; DB 2; Length 398;
Best Local Similarity 53.8%; Pred. No. 5.4e-42;
Matches 233; Conservative 56; Mismatches 108; Indels 36; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSSATLSAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTTQ 60
Db 1 MKKRILSAVLVSGVTLGAATT---VGAEDLSKIAKQDSIISNLTTEKQAQNVSAQ 57
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQAQSNAA 120
Db 58 QVSSLSQEQDKLTARNTLEALSKEFEIKALTSQIVARNEKLNQARSAYKNNETSGY 117
QY 121 INAIINSKVSDAINRVSAIRVSVANEKMLHQEQDKAAVEQKHQENQAINTVAAN 180
Db 118 INALLNSKISDVNRLVAINRAVSANAKLLEQKADKVSLEEKQAQNTAINTIAAN 177
QY 181 TIAQNTNALNTQOALEAAQLNLOAELTTAODQKATLVAKAAAEAAEAQAAAAA 240
Db 178 MAEENQNTLRTQANLEAATAANLALQLASATEDKANLVAKAEAAEAALAQEA 237
QY 241 KAAAEKALQEAQAQAAQAAANNNTQATDASDQQAADNTQAAQTGSDTDSQA 300
Db 238 KA-----QEQA-----QQAASVEAKSAITPAPQATPAAQSSNA 272
QY 301 SDQESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPVGNWNGG 360
Db 273 IEPALTAAPAA-PSARPOT-----SYDSSNTYPVQCTGWGKSLAPWAGN 325
QY 361 ASAAAAGRVGTPSAGAVAVNWDGQYGHVAVTVGVQ--GGQIQVQEAANYAG 419
Db 326 YSAQAAGVTRGTPMVGAIVNWDGQYGHVAVTVGVQSSASSIRVMESYGRQY 385
QY 420 WFNPGSVSYIYPN 432
Db 386 WFNPTGVTFIYPH 398

RESULT 13
Q8P318 STRP3 PRELIMINARY; PRT; 398 AA.
AC Q8P318; Q7CFL7;
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 13-SEP-2005 (TremBLrel. 31, Last annotation update)
DE Putative secreted protein.
GN OrderedLocusNames=SP0015; Spym3_0014;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301448;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SSI-1 / Serotype M3;
RX MEDLINE=22683278; PubMed=12799345; DOI=10.1101/gr.1096703;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution."
RL Genome Res. 13:1042-1055(2003).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MGAS315 / Serotype M3;
```



RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
RN [2]
RN NUCLEOTIDE SEQUENCE.
RP Fagan P.K., Reinscheid D., Gottschalk B., Chhatwal G.S.;
RT "Identification and characterization of a novel secreted protein from
RT group A streptococcus.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE006474; AAK3158.1; -; Genomic DNA.
DR EMBL; AF319999; AAL73135.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KT Complete proteome; Signal.
FT SIGNAL 1 23 Potential.
SQ SEQUENCE 398 AA; 41899 MW; 28A9B3F7195E969B CRC64;
Query Match 51.1%; Score 1081; DB 2; Length 398;
Best Local Similarity 53.6%; Pred. No. 1e-41;
Matches 232; Conservative 56; Mismatches 109; Indels 36; Gaps 6;
QY 1 MKKRILSAVLVSGVTLSSATLTSVAKADPDFDAQIASODSKINNLTAAQQAAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSSATLTSVAKADPDFDAQIASODSKINNLTAAQQAAQAQVNTIQG 57
QY 61 QVSALQTOQAELEQAEQNRLEAQSATLGGQIQITLSSXIVARNESLKQOARSQAQSNAAATSY 120
DB 58 QVSSLSQSEQDKLTARTELEALSKEFEQIKALTSQIVARNKLNQARSAYKNNETSGY 117
QY 121 INAINSKSVDAINRVSAIREVVSVANERKMLHQEQODKAAVROKHQENQAAINTVAANOE 180
DB 118 INALLNSKISIDVYVRLVAIRAVNSANAKLLEQOKADKVSLEEKQAAQNTAINTTAAANMA 177
QY 181 TTAQNTNALNTQOQLERAAQLNLQALTLTAQDQKATLVAQKAAAEAEARQAAQAQAAEA 240
DB 178 MAEENQTLRTQOANLVAATANLALQLASATEDKANLVAQKAAAEKAAAEALAEQAAKV 237
QY 241 KAAAEAKALQEAQAQAQAANNNNTQATDASDQQAADNTQAAQTGSDTQSAQAQVNN 300
DB 238 KA-----QEQA-----QQAASVEAKSAITPAQATPAQSSNA 272
QY 301 SDOESTTATAQPSASSASTAAVAANTSSANTYPAGCCTGWGVKSLAPWVGNVWNGGQWA 360
DB 273 IEPAALTAPAA-PSAGPQT-----SYDSNTYVPVGCCTWGAKSLAPWAGNNWNGGQWA 325
QY 361 ASAAAGYRVGTPSPAGAVVNDGGYGHVAVYTVQO-GGQIQVQENYAGNQSIGNYEG 419
DB 326 YSAQAAGYRTGSTPMVGAIVAVNDGGYGHVAVVVEVQSASSIRVMESNYSGRQYTADHRG 385
QY 420 WFNPGSVSYIYPN 432
DB 386 WFNPTGVTFIYPH 398

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:09:37 ; Search time 30.0232 Seconds  
(without alignments)  
1189.611 Million cell updates/sec

Title: US-10-797-821-32  
Perfect score: 2116  
Sequence: 1 MKKRLSAVLVSGVTLSAT.....SIGNRYGNFPGSVYIYPN 432

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5 COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/6 COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/H COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/pCTUS COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/RE COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description          |
|------------|-------|-------------|--------|-------|----------------------|
| 1          | 1053  | 49.8        | 399    | 2     | US-09-107-433-3230   |
| 2          | 1052  | 49.7        | 392    | 2     | US-09-583-110-4374   |
| 3          | 679.5 | 32.1        | 461    | 1     | US-08-186-222-2      |
| 4          | 581.5 | 27.5        | 525    | 2     | US-09-107-532A-5095  |
| 5          | 555   | 26.2        | 210    | 2     | US-09-222-938A-67    |
| 6          | 527   | 24.9        | 497    | 2     | US-09-134-000C-5990  |
| 7          | 494.5 | 23.4        | 449    | 2     | US-09-071-035-482    |
| 8          | 494.5 | 23.4        | 449    | 2     | US-10-206-576-482    |
| 9          | 494.5 | 23.4        | 450    | 2     | US-09-134-000C-5714  |
| 10         | 445.5 | 21.1        | 422    | 2     | US-09-071-035-484    |
| 11         | 445.5 | 21.1        | 422    | 2     | US-10-206-576-484    |
| 12         | 280.5 | 13.3        | 469    | 2     | US-09-489-039A-13565 |
| 13         | 237   | 11.2        | 264    | 2     | US-09-134-001C-5035  |
| 14         | 236   | 11.2        | 257    | 2     | US-09-710-279-3244   |
| 15         | 236   | 11.2        | 267    | 2     | US-09-134-001C-4539  |
| 16         | 230   | 10.9        | 477    | 2     | US-09-902-540-11649  |
| 17         | 230   | 10.9        | 2310   | 2     | US-08-874-923-120    |
| 18         | 224   | 10.6        | 1236   | 2     | US-09-769-787-109    |
| 19         | 222.5 | 10.5        | 270    | 2     | US-09-134-001C-5441  |
| 20         | 214.5 | 10.1        | 266    | 2     | US-09-134-001C-5453  |
| 21         | 214   | 10.1        | 149    | 2     | US-09-710-279-1682   |
| 22         | 214   | 10.1        | 157    | 2     | US-09-710-279-2870   |
| 23         | 213   | 10.1        | 610    | 2     | US-09-336-447A-11    |
| 24         | 213   | 10.1        | 610    | 2     | US-09-952-267B-11    |
| 25         | 212.5 | 10.0        | 468    | 2     | US-09-328-352-6321   |
| 26         | 211.5 | 10.0        | 639    | 2     | US-09-902-540-14908  |
| 27         | 211.5 | 10.0        | 655    | 2     | US-09-902-540-10005  |

|    |       |      |      |   |                      |                   |
|----|-------|------|------|---|----------------------|-------------------|
| 28 | 211   | 10.0 | 440  | 2 | US-08-302-756E-35    | Sequence 35, Appl |
| 29 | 206.5 | 9.8  | 1566 | 1 | US-08-687-956A-23    | Sequence 23, Appl |
| 30 | 206   | 9.7  | 679  | 2 | US-09-489-039A-12307 | Sequence 12307, A |
| 31 | 201.5 | 9.5  | 955  | 1 | US-08-006-676B-1     | Sequence 1, Appl1 |
| 32 | 201.5 | 9.5  | 955  | 1 | US-08-282-845-2      | Sequence 2, Appl1 |
| 33 | 201.5 | 9.5  | 955  | 1 | US-08-428-414A-3     | Sequence 3, Appl1 |
| 34 | 201.5 | 9.5  | 955  | 4 | PCT-US94-00324-1     | Sequence 1, Appl1 |
| 35 | 201   | 9.5  | 1070 | 2 | US-09-902-540-13861  | Sequence 13861, A |
| 36 | 195   | 9.2  | 2107 | 2 | US-09-949-016-7646   | Sequence 7646, Ap |
| 37 | 195   | 9.2  | 2107 | 2 | US-09-949-016-7647   | Sequence 7647, Ap |
| 38 | 195   | 9.2  | 2115 | 2 | US-09-296-662-33     | Sequence 33, Appl |
| 39 | 194.5 | 9.2  | 883  | 2 | US-09-489-039A-12755 | Sequence 12755, A |
| 40 | 193.5 | 9.1  | 2101 | 1 | US-08-466-390-4      | Sequence 4, Appl1 |
| 41 | 193.5 | 9.1  | 2101 | 1 | US-08-470-950-4      | Sequence 4, Appl1 |
| 42 | 193.5 | 9.1  | 2101 | 1 | US-08-467-781-4      | Sequence 4, Appl1 |
| 43 | 193.5 | 9.1  | 2101 | 1 | US-08-195-487-4      | Sequence 4, Appl1 |
| 44 | 193.5 | 9.1  | 2101 | 1 | US-08-483-924-4      | Sequence 4, Appl1 |
| 45 | 193.5 | 9.1  | 2101 | 2 | US-09-452-294-1      | Sequence 1, Appl1 |

ALIGNMENTS

RESULT 1

US-09-107-433-3230  
; Sequence 3230, Application US/09107433  
; Patent No. 6800744

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE

NUMBER OF SEQUENCES: 5206

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: <Unknown>  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: <Unknown>

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,433  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085131  
FILING DATE: May 12, 1998

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneka  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 3230:

SEQUENCE CHARACTERISTICS:  
LENGTH: 399 amino acids  
TYPE: amino acid

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Streptococcus pneumoniae

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...399

SEQUENCE DESCRIPTION: SEQ ID NO: 3230:

US-09-107-433-3230

Query Match 49.8%; Score 1053; DB 2; Length 399;  
Best Local Similarity 49.8%; Pred. No. 2.3e-73;  
Matches 217; Conservative 74; Mismatches 93; Indels 52; Gaps 4;

Qy 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAQOQAAQAVNTIQG 60  
Db 8 MKKKILASILLSTVMSVQAVLTTAAHTTDDKIAAQDNKISNLTAQOQAAQAVNTIQG 67

Qy 61 QVSALQTOQAEIQAEORLEAQSAITLGOQIOTLSSKIVARNESLKKQARSQAQKSNAAITSY 120  
Db 68 QVSAIQAEQSNLQAEENDRLQAEKKEGTEITELSKNIVSRNQSLKKQARSQAQKSNAAITSY 127

Qy 121 INALINSKVSDAINRVSASIREVVSANKEKMLHQEQDQKAAVEQKHQENQAINTVAANOE 180  
Db 128 INTIVNSKSITEAISRVAAEMSEIVSANNKMLEQQKADKKAISEKQVANNDAINTVIANQQ 187

Qy 181 TIAQNTNALNTQOQAEQLEAAQNLQAEELTTAQDQKATLVAQKAAAEAEARQAAAAQAAAA 240  
Db 188 KLADDAQALTTKQAEKKAELSAAEKATAGEKASLEQKAAAEAEARQAAAAQAAAA 247

Qy 241 KAAAEKALQOQAAQAAAAANNNTQATDASDQOQAAAAADNTQAAQTGSDTQDQAAAAVNN 300  
Db 248 KRASQOQSV-----LASANTNLTAQVQAVSESAAAAAPRAKVRPT----- 286

Qy 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPVGVNMGNGGOWA 360  
Db 287 -----YSTNASSYPIGECTGWGKSLAPVGVNMGNGGOWA 321

Qy 361 ASAAAAGYRVGSTPSAGAVAVNNDGGYGHVAVVTGVQ- GOIQVQEAANYAGNOSIGNVRG 419  
Db 322 TSAAAAGPRTGSTPQVGAIAACWNDGGYGHVAVVTAVESTTRIQVSESANYAGNRTIGNHRG 381

Qy 420 WFNPN-----GSVSYIY 430  
Db 382 WFNPTTTSEGFVTIY 397

RESULT 2

US-09-583-110-4374

; Sequence 4374, Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE REFERENCE: PATH00-07A  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 4374  
; LENGTH: 392  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae

US-09-583-110-4374

Query Match 49.7%; Score 1052; DB 2; Length 392;  
Best Local Similarity 49.8%; Pred. No. 2.6e-73;  
Matches 217; Conservative 73; Mismatches 94; Indels 52; Gaps 4;

Qy 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAQOQAAQAVNTIQG 60  
Db 1 MKKKILASILLSTVMSVQAVLTTAAHTTDDKIAAQDNKISNLTAQOQAAQAVNTIQG 60

Qy 61 QVSALQTOQAEIQAEORLEAQSAITLGOQIOTLSSKIVARNESLKKQARSQAQKSNAAITSY 120

Db 61 QVSAIQAEQSNLQAEENDRLQAEKKEGTEITELSKNIVSRNQSLKKQARSQAQKSNAAITSY 120

Qy 121 INALINSKVSDAINRVSASIREVVSANKEKMLHQEQDQKAAVEQKHQENQAINTVAANOE 180  
Db 121 INTIVNSKSITEAISRVAAEMSEIVSANNKMLEQQKADKKAISEKQVANNDAINTVIANQQ 180

Qy 181 TIAQNTNALNTQOQAEQLEAAQNLQAEELTTAQDQKATLVAQKAAAEAEARQAAAAQAAAA 240  
Db 181 KLADDAQALTTKQAEKKAELSAAEKATAGEKASLEQKAAAEAEARQAAAAQAAAA 240

Qy 241 KAAAEKALQOQAAQAAAAANNNTQATDASDQOQAAAAADNTQAAQTGSDTQDQAAAAVNN 300  
Db 241 KRASQOQSV-----LASANTNLTAQVQAVSESAAAAAPRAKVRPT----- 279

Qy 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPVGVNMGNGGOWA 360  
Db 280 -----YSTNASSYPIGECTGWGKSLAPVGVNMGNGGOWA 314

Qy 361 ASAAAAGYRVGSTPSAGAVAVNNDGGYGHVAVVTGVQ- GOIQVQEAANYAGNOSIGNVRG 419  
Db 315 TSAAAAGPRTGSTPQVGAIAACWNDGGYGHVAVVTAVESTTRIQVSESANYAGNRTIGNHRG 374

Qy 420 WFNPN-----GSVSYIY 430  
Db 375 WFNPTTTSEGFVTIY 390

RESULT 3

US-08-186-222-2

; Sequence 2, Application US/08186222  
; Patent No. 5559007  
; GENERAL INFORMATION:  
; APPLICANT: Suri, Bruno  
; APPLICANT: Schmitz, Albert  
; TITLE OF INVENTION: Bacterial Vectors  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/186,222  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/672,205  
; FILING DATE: 19-MAR-1991  
; APPLICATION NUMBER: GB 9006400.7  
; FILING DATE: 22-MAR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Villamizar, JoAnn  
; REGISTRATION NUMBER: 30,598  
; REFERENCE/DOCKET NUMBER: 4-17994/A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (914)785-7121  
; TELEFAX: (914)347-5769  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 461 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-186-222-2

Query Match 32.1%; Score 679.5; DB 1; Length 461;



```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 67
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-222-938A-67

Query Match
Best Local Similarity 26.2%; Score 555; DB 2; Length 210;
Matches 115; Conservative 44; Mismatches 51; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
Db 1 MKKKILASLLSTVMVSQVAVLTTHAETDDKIAAQDNKISNLTAQQAQAQVNDIQE 60

QY 61 QVSALQTOQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKKQARSQAQKNAATSY 120
Db 61 QVSAIQAEQNSLQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKKQARSQAQKNAATSY 120

QY 121 INAIINSKSVSDAIRVSAIREVVSANKEKMLHQEQDQKAAVEQKHQENQAANTVAANOE 180
Db 121 INAIINSKSVSDAIRVSAIREVVSANKEKMLHQEQDQKAAVEQKHQENQAANTVAANOE 180

QY 181 TTAQNTNALNTQQAQLAQLNLQAEELTTA 210
Db 181 KLADDAQAALTQKQELKAAELSLAAEKATS 210

RESULT 6
US-09-134-000C-5990
; Sequence 5990, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5990
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5990

Query Match
Best Local Similarity 24.9%; Score 527; DB 2; Length 497;
Matches 152; Conservative 82; Mismatches 175; Indels 76; Gaps 12;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
Db 16 LKSVLSALMVCSTILTSVALPSAFADYDFKIQQDQKINALTSQMSDAEAKVAIEN 75

QY 61 QVSALQTOQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKKQARSQAQKNAATSY 120
Db 76 DMVETAKQIDTLTAKNKLSSVSKLYSEISDLNVRIOKREVQMTKQARDVQVNGQSDSI 135

QY 121 INAIINSKSVSDAIRVSAIREVVSANKEKMLHQEQDQKAAVEQKHQENQAANTVAANOE 180
Db 136 IDAVIDADSDAIDIRGQAVSTMSSANNLELQQEQDQKATVEKTKNVKEQIAELEAATK 195

QY 181 TTAQNTNALNTQQAQLAQLNLQAEELTTA 210
Db 196 ELNDKTESLTKLQIQEVAKNLDEAQRSEQKDGFKQKKEAEKRLAEQARQAAAK 255

QY 235 -----QAAAEAKAAAEAKALQEQQAQAQAANNNTNTQATDASQQAQAANTQAAQTGST 290
Db 235 -----QAAEQAAAQAQAQAQ- KAAAEQAKATKAA-----NEAAASAAAEKAA-----TPVVSSTTT 306
```

```
QY 291 DOSAAQAVNNSDQES-----TTATAAQP-----SASSASTA 321
Db 307 ESTTQETTTSTETESVVTTPVAAAPEKEKEVPVNTTPEKGNKAPGNGGVTSKQA 366

QY 322 AV-AAANTSSANTYPA-----GQCTWGVKSLAPWVGVNGWGQWMAASAAAAGYRVG-STP 374
Db 367 AINAAALADVGNYSATGMNQPGCECLVSVRRWLAAGINFGYGGPNSGYVASGATQVSWSNV 426

QY 375 SAGAVAVNDG-----GYGHVAVVTGQGGQIQVQEAANYAGNSIQINRYRWFNPGSVS 427
Db 427 QPGDVVQVESAYSPDSWTGGVHTVLVTGVSQSVQIVAN-----NPGSG 472

QY 428 YIYPN 432
Db 473 YVSSN 477

RESULT 7
US-09-071-035-482
; Sequence 482, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 482:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-482

Query Match
Best Local Similarity 23.4%; Score 494.5; DB 2; Length 449;
Matches 138; Conservative 87; Mismatches 177; Indels 51; Gaps 12;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
Db 1 VKKRILFASVLLCSLTLSAIAATPSIALADNVDKIEKNQEISSLKAKQGDLSAQVSSLEA 60

QY 61 QVSALQTOQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKKQARSQAQKNAATSY 120
Db 61 EVSSVFDSESMALREKQKTLKARSEQLOQEITNLNRIEKNREAIKNQARDVQVNGQSTTM 120

QY 121 INAIINSKSVSDAIRVSAIREVVSANKEKMLHQEQDQKAAVEQKHQENQAANTVAANOE 180
```



Db 182 ELETQRDQLLSQSEINLVKASLALAEQSSAESKAGLEKQKAAAEQARLAAEQKAAE 241  
QY 227 AARQAAAAQAAAEAKAAAEAKALQEQAAQAAANNNNTQATDASDQAAA-----AADNT 281  
Db 242 KAKQAAAKPAKAEVK--AEAPVASSSTTEAQAPASSSS--ATESSTQQTETTTTPSDTNS 297  
QY 282 QAAQCGDSTDQAAQAV-----NNSDQESTTATAAQPSASSASTAAVAANTSSANTYPA 335  
Db 298 ATENTGSSSSSEPOVPTTPTSDNGNGGQGTGTTPTPTPAPSADPTINALNVLROSL 357  
QY 336 GCGTWGKSLAPWNGYNGWGQOWAASAAAGYRVGSTPSAGAVAVWMDGGYGHVAYVTG 395  
Db 358 G-----LRPVV---WDAGLAASATARAQVEAGGIPNDH-----WSRG--DEVIAIMW 400  
QY 396 VQGGQIQV---QEANYAGNQSIGNYRGW-FNPG 424  
Db 401 AFGNSVIMAWYNETNMTASGSG-HRDWEINPG 432  
RESULT 10  
US-09-071-035-484  
; Sequence 484, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB369P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 484:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 422 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-071-035-484  
Query Match 21.1%; Score 445.5; DB 2; Length 422;  
Best Local Similarity 29.8%; Pred. No. 1.9e-26;  
Matches 127; Conservative 80; Mismatches 168; Indels 51; Gaps 12;  
QY 28 DFDQAIASQDSKINNLTAAQQAQAAQVNTIQGYVSALQTQQAELQAEORLEAQASATLG 87  
Db 1 DNVDKKEIKQKESLSLKAKQGDLSAQVSSLEAEVSSVFDSEMLREKQTLKASEQLQ 60  
QY 88 QQIOTLSSKIVARNESLKKQQAASQAQSNAAATSYINAINIINSKSVDAINRVSAIREVWSAN 147

Db 61 QBITNLQRIEKRNAIKQARDVQVNGQSTMTLMDAVIDADSVADAISRQVQAVSTIVSAN 120  
QY 148 EKMLHQODKDAAYVEQKHQENQAAINTVAANQETIAQNTNALNTQQAQLAAQLNLAEL 207  
Db 121 NDLMOQOKEDQAVVDKKAENEKKVQKLEATEAELETKRQDILLSKSELNVKASLAEQ 180  
QY 208 TTAODQKATLVAQKAAAE-----EAAQAAAAQAAAAEAKAAAEAKALQEOA 253  
Db 181 SSASSKAGLEKQKAAAEQARLAAEQKAAAEKAKQAAAKPAKAEVK--AEAPVASSST 238  
QY 254 AQAQAAANNNNTQATDASDQAAA-----AADNTQAAQTGDDSTDSAAQAV-----NNSD 302  
Db 239 TEAQAPASSSS--ATESSTQQTETTTTPSDTNSATENTGSSSSSEPOVPTTPTSDNGNGG 296  
QY 303 QESTTATAAQPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWNGYNGWGQOWAAS 362  
Db 297 QTGGGTVTPTPTPAPSADPTINALNVLROSLG-----LRPVV---WDAGLAASAT 345  
QY 363 AAAGYRVGSTPSAGAVAVWMDGGYGHVAYVTGVOGGQIQV---QEANYAGNQSIGNYRG 419  
Db 346 ARAAQVEAGGIPNDH-----WSRG--DEVIAIMWAFGNSVIMAWYNETNMTASGSG-HRD 398  
QY 420 W-FNPG 424  
Db 399 WEINPG 404  
RESULT 11  
US-10-206-576-484  
; Sequence 484, Application US/10206576  
; Patent No. 6913907  
; GENERAL INFORMATION:  
; APPLICANT: Choi et al.  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 497  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-R  
; COMPUTER: Dell Latitude  
; OPERATING SYSTEM: Windows 98  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/206,576  
; FILING DATE: 29-Jul-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/071,035  
; FILING DATE: 1998-05-04  
; APPLICATION NUMBER: US 60/046,655  
; FILING DATE: 1997-05-16  
; APPLICATION NUMBER: US 60/044,031  
; FILING DATE: 1997-05-06  
; APPLICATION NUMBER: US 60/066,009  
; FILING DATE: 1997-11-14  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hyman, Mark J.  
; REGISTRATION NUMBER: 46,789  
; REFERENCE/DOCKET NUMBER: PB369PID1  
; INFORMATION FOR SEQ ID NO: 484:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 422 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 484:  
US-10-206-576-484

Query Match 21.1%; Score 445.5; DB 2; Length 422;  
Best Local Similarity 29.8%; Pred. No. 1.9e-26;  
Matches 127; Conservative 80; Mismatches 168; Indels 51; Gaps 12;

QY 28 DFDIAQIASQSKINLTAQQAQAQAVNTTIGQVSAQTQQAELQAEINQRLEAQSATIG 87  
DB 1 DNVDKIEKQEISSLKAKQGDLSQVSSLEAEYSVFDSEMLREQKQTLKAKSEQIQ 60  
QY 88 QOIQTLSSKIVARNESLQQAARSQAQSNAAATSYINAININSKVSDAINRVSAIREVGSAN 147  
DB 61 QBITVNLORIEKRNIAIKNQARDVQVNGQSTTMDAVIDLSDVADISRVQAVSTIVSAN 120  
QY 148 EKMLHQEQBDKAAVQKQHQENQAAINTVAANQETIAQNTNALNTQQAQLEAAQLNLOAEL 207  
DB 121 NDLMQOQKEDQAVVDKKAENEKKVKQLEATEAELETKRQDLSKQSELNVKMSLASLEQ 180  
QY 208 TTAQOKATLVAKAAAE-----EAAQAAAAQAAAEAKAAAEAKALQEOA 253  
DB 181 SSAESKAGLEKQKAAAEAEQARLAAEQKAAAEKAKQAAKPAKAEVK--AEAPVASST 238  
QY 254 AQAQAANNNTQATDASDQAAA-----AADNTQAAQTGSDTDQSAQAQAV-----NNSD 302  
DB 239 TEAQAPASSS--ATESSTQQTETTTPTSDNTSENTGSSSEBPQVPTTSPDNGNGG 296  
QY 303 QESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWGVNGVGGQWAAAS 362  
DB 297 QTGGGTVTPTPTPAPSADPTINALNVLRSLSG-----LRPVV---WDAGLAASAT 345  
QY 363 AAAAGYRGVSTPSAGAVAVNDGGGHVAVYTVGVGGQIQV---QEAQVAGNQSIGNYRG 419  
DB 346 ARAAQVEAGGIENDH-----WSRG--DEVIAIMWAFGNSVIMAWYNETNMVTASGSG-HRD 398  
QY 420 W-FNPG 424  
DB 399 WEINPG 404

RESULT 12  
US-09-489-039A-13565  
; Sequence 13565, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 13565  
; LENGTH: 469  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-13565

Query Match 13.3%; Score 280.5; DB 2; Length 469;  
Best Local Similarity 28.9%; Pred. No. 1.2e-13;  
Matches 113; Conservative 56; Mismatches 165; Indels 57; Gaps 12;

QY 31 DAQIASQSKINLTAQQAQ-----AAQAVNTTIGQVSAQTQQAELQAEINQRLEAQSATIG 78  
DB 84 DAIMVDPGAVNNVNRQQAQASARAEQREKQQAQAELEKQAAEQERLQLEQER 143  
QY 79 LEAQSATLGGQQTQLSSKIVARNESLQQAARSQAQSNAAATSYINAININSKVSDAINRV 138  
DB 144 LQAQEAQAEKQ-----QKQAEAAKAAAKAAKAKADAQAKAEQAEAAKAA 190  
QY 139 AIREVVSAENKMLHQEQBDKAAVEQKHQENQAAINTVA-ANQETIAQNTNALNTQQAQLE 197  
DB 191 AEAKAKADAQAKAEQAAKAAADAKQAAEAAKAAAEAKKQAAEAAKAAAEAKKAAE 250

QY 198 AA---QLNLOAELTTAQDQKATLVAKAAAEAEARQAAAAQAAAAQAAAAEAKAAAEAKALQEOAA 254  
DB 251 AAAAKKAQQAQAEKKAQQAQAAAAEAKAAAEAKAAAEAKAAAEAKAAAEAKAAAEAKAAAEAKAA 309  
QY 255 QAAAAANNNTQATDASDQQAQAADNTQAAQTGSDTD-----QSAQAQVNNSDQES 305  
DB 310 ABEKAAAD---KAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 366  
QY 306 TTATAQPSASSASTA-----AAVAANTSSANTYPAGQCTWGVKSLAP--WVGNY 352  
DB 367 AAAAGSGMTKNSASGADINNVAGQIKSAIESKFYDASSYAGTKTCTLRK-LAPDGLLLNI 425  
QY 353 WNGCGOWA-ASAAAAAGYRGVSTPSAGAVAVW 382  
DB 426 QSEGGDPALCQAALAAARQAQKFPKPPSQAVY 456

RESULT 13  
US-09-134-001C-5035  
; Sequence 5035, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5035  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5035

Query Match 11.2%; Score 237; DB 2; Length 264;  
Best Local Similarity 34.7%; Pred. No. 1.3e-10;  
Matches 60; Conservative 27; Mismatches 70; Indels 16; Gaps 5;

QY 260 ANNNNTQATDASDQQAQAADNTQAAQTGSDTDQSAQAQVNNSDQESTTTATAAQPSSASSAS 319  
DB 104 SNYNNYQ-----SNNTQSORTTQPTGGLGASYSTSSNVHVTTTSA-PSSNGVS 151  
QY 320 TAAVAANTSSANTYPAGQCTWGV-KSLAPWGVNNGVGGQWAAASAAAAAGYRGVSTPSAGA 378  
DB 152 LS--NARSAGNLYTSGQCTYVYFDRVGKIGSTWGNANNWANAARSGYTYNNSPAKGA 209  
QY 379 VAVWMDGCGYGHVAVYTVGV-QGQIQVQEAQVAGNOSIGNYRGWFPNGSVSYIY 430  
DB 210 ILQTSQGAIGHVAVYVEGVNSNGSIRVSENNYGHGAGVTVTSTISASQAASNY 262

RESULT 14  
US-09-710-279-3244  
; Sequence 3244, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3244  
; LENGTH: 257





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:29:27 ; Search time 92.071 Seconds  
(without alignments)  
1960.467 Million cell updates/sec

Title: US-10-797-821-32

Perfect score: 2116

Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNRWGFWNPGSVSIYPN 432

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA Main:\*
- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
  - 3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
  - 6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID              | Description       |
|------------|--------|-------------|--------|--------------------|-------------------|
| 1          | 2116   | 100.0       | 432    | US-10-383-930-32   | Sequence 32, Appl |
| 2          | 2116   | 100.0       | 432    | US-10-797-821-32   | Sequence 32, Appl |
| 3          | 2099   | 99.2        | 432    | US-10-383-930-31   | Sequence 31, Appl |
| 4          | 2099   | 99.2        | 432    | US-10-797-821-31   | Sequence 31, Appl |
| 5          | 2074.5 | 98.0        | 431    | US-10-383-930-30   | Sequence 30, Appl |
| 6          | 2074.5 | 98.0        | 431    | US-10-797-821-30   | Sequence 30, Appl |
| 7          | 2070.5 | 97.8        | 431    | US-10-383-930-29   | Sequence 29, Appl |
| 8          | 2070.5 | 97.8        | 431    | US-10-797-821-29   | Sequence 29, Appl |
| 9          | 2066.5 | 97.7        | 431    | US-10-383-930-33   | Sequence 33, Appl |
| 10         | 2066.5 | 97.7        | 431    | US-10-797-821-33   | Sequence 33, Appl |
| 11         | 1081   | 51.1        | 398    | US-10-474-792-600  | Sequence 600, App |
| 12         | 1053   | 49.8        | 392    | US-10-472-928-4652 | Sequence 4652, Ap |
| 13         | 1053   | 49.8        | 399    | US-10-617-320-3230 | Sequence 3230, Ap |
| 14         | 585.5  | 27.7        | 524    | US-10-797-821-67   | Sequence 6758, A  |
| 15         | 555    | 26.2        | 210    | US-10-154-251-67   | Sequence 6758, A  |
| 16         | 494.5  | 23.4        | 449    | US-09-071-035-482  | Sequence 482, App |
| 17         | 494.5  | 23.4        | 449    | US-10-206-576-482  | Sequence 482, App |
| 18         | 494.5  | 23.4        | 449    | US-10-912-362-482  | Sequence 482, App |
| 19         | 445.5  | 21.1        | 422    | US-09-071-035-484  | Sequence 484, App |
| 20         | 445.5  | 21.1        | 422    | US-10-206-576-484  | Sequence 484, App |
| 21         | 445.5  | 21.1        | 422    | US-10-912-362-484  | Sequence 484, App |
| 22         | 320    | 15.1        | 630    | US-10-494-674-6    | Sequence 6, Appli |
| 23         | 314    | 14.8        | 600    | US-09-738-626-5197 | Sequence 5197, Ap |
| 24         | 300.5  | 14.2        | 609    | US-10-501-282-3184 | Sequence 3184, Ap |
| 25         | 293.5  | 13.9        | 440    | US-10-501-282-3056 | Sequence 3056, Ap |
| 26         | 291    | 13.8        | 422    | US-10-501-282-3054 | Sequence 3054, Ap |
| 27         | 282    | 13.3        | 377    | US-10-501-282-3052 | Sequence 3052, Ap |

ALIGNMENTS

RESULT 1

US-10-383-930-32

; Sequence 32, Application US/10383930

; Publication No. US20040127400A1

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J

; APPLICANT: Taubman, Martin A

; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein

; FILE REFERENCE: 25669-018

; CURRENT APPLICATION NUMBER: US/10/383,930

; CURRENT FILING DATE: 2003-03-07

; PRIOR APPLICATION NUMBER: 60/402,483

; PRIOR FILING DATE: 2002-08-08

; PRIOR APPLICATION NUMBER: 60/363,209

; PRIOR FILING DATE: 2002-03-07

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 32

; LENGTH: 432

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-10-383-930-32

Query Match 100.0%; Score 2116; DB 4; Length 432;

Best Local Similarity 100.0%; Pred. No. 1.9e-121; Mismatches 0; Indels 0; Gaps 0;

Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKKRILSAVLVSGVTLSATLSAVKADDDFAQIASQDSKINNLTAAQQAQAQVNTIQG 60

Db 1 MKKRILSAVLVSGVTLSATLSAVKADDDFAQIASQDSKINNLTAAQQAQAQVNTIQG 60

Qy 61 QVSALQTOQAELOAENORLEAQSAITLGGQIQTLSSKIVARNESLKQARSQAQKSAATS 120

Db 61 QVSALQTOQAELOAENORLEAQSAITLGGQIQTLSSKIVARNESLKQARSQAQKSAATS 120

Qy 121 INAIINSKVSADINRVSAIREVVSANESKMLHQQDQKAAVEQKHQENQAALNTVAANOE 180

Db 121 INAIINSKVSADINRVSAIREVVSANESKMLHQQDQKAAVEQKHQENQAALNTVAANOE 180

Qy 181 TIAQNTNALNTQQAQLEAAQLNLQAEITTAQDQKATLVAQKAAAEAAARQAQAAQAAEA 240

Db 181 TIAQNTNALNTQQAQLEAAQLNLQAEITTAQDQKATLVAQKAAAEAAARQAQAAQAAEA 240

Qy 241 KAAAEAKALQEQAAQAAQAAANNNNTQATDASDQAAAADNTQAAQTGSDTDSAAQAVNN 300

Db 241 KAAAEAKALQEQAAQAAQAAANNNNTQATDASDQAAAADNTQAAQTGSDTDSAAQAVNN 300

Qy 301 SDOESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPVGNVWNGGGOWA 360

Db 301 SDOESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPVGNVWNGGGOWA 360



```

; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-31

Query Match          99.2%; Score 2099; DB 5; Length 432;
Best Local Similarity 99.3%; Pred. No. 2.1e-120;
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60
Db 1 MKKRILSAVLVSGVTLSATTLSAIAKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60

QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQITLSSKIVARNESLKQOARSAQKSNAAATSY 120
Db 61 QVSALQTOQAELOAENQRLAQSATLGGQIQITLSSKIVARNESLKQOARSAQKSNAAATSY 120

QY 121 INAIINSKSVSDAIRVSAIREVWSANEKMLHQODKAAVEQKHQENQAAINTVAANQE 180
Db 121 INAIINSKSVSDAIRVSAIREVWSANEKMLHQODKAAVEQKHQENQAAINTVAANQE 180

QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAODQKATLVAKAAAEAEARQAAAAQAQAAEA 240
Db 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAODQKATLVAKAAAEAEARQAAAAQAQAAEA 240

QY 241 KAAAEAKALQEQAQAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAQVNN 300
Db 241 KAAAEAKALQEQAQAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAQVNN 300

QY 301 SDOESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVNTWNGGGQWA 360
Db 301 SDOESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVNTWNGGGQWA 360

QY 361 ASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
Db 361 ASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 420

QY 421 FNPGSVSIYYPN 432
Db 421 FNPGSVSIYYPN 432

RESULT 5
US-10-383-930-30
; Sequence 30, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08

; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-30

Query Match          98.0%; Score 2074.5; DB 4; Length 431;
Best Local Similarity 98.8%; Pred. No. 6.6e-119;
Matches 427; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60

QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQITLSSKIVARNESLKQOARSAQKSNAAATSY 120
Db 61 QVSALQTOQAELOAENQRLAQSATLGGQIQITLSSKIVARNESLKQOARSAQKSNAAATSY 120

QY 121 INAIINSKSVSDAIRVSAIREVWSANEKMLHQODKAAVEQKHQENQAAINTVAANQE 180
Db 121 INAIINSKSVSDAIRVSAIREVWSANEKMLHQODKAAVEQKHQENQAAINTVAANQE 180

QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAODQKATLVAKAAAEAEARQAAAAQAQAAEA 240
Db 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAODQKATLVAKAAAEAEARQAAAAQAQAAEA 240

QY 241 KAAAEAKALQEQAQAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAQVNN 300
Db 241 KAAAEAKALQEQAQAQAAANNNTQATDASDQAAAAADNTQAAQTGSDTDSAAQAQVNN 300

QY 301 SDOESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVNTWNGGGQWA 360
Db 301 SDOESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVNTWNGGGQWA 360

QY 361 ASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 420
Db 361 ASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYVTGVGGQIQVQEAANYAGNOSIGNYRGW 420

QY 421 FNPGSVSIYYPN 432
Db 421 FNPGSVSIYYPN 432

RESULT 6
US-10-797-821-30
; Sequence 30, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
```

; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-821-30

Query Match 98.0%; Score 2074.5; DB 5; Length 431;  
Best Local Similarity 98.8%; Pred. No. 6.6e-119;  
Matches 427; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTTIG 60  
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTTIG 60  
Qy 61 QVSALOTQOAELOAENORLEAQSATLGGQIOTLSSKI VARNESLKQOARSQAQKSNAAATSY 120  
Db 61 QVSALOTQOAELOAENORLEAQSATLGGQIOTLSSKI VARNESLKQOARSQAQKSNAAATSY 120  
Qy 121 INAIINSKSVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENOAINTVAANQE 180  
Db 121 INAIINSKSVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENOAINTVAANQE 180  
Qy 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAODOKATLVAQKAAAEAEARQAAAAQAAEA 240  
Db 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAODOKATLVAQKAAAEAEARQAAAAQAAEA 240  
Qy 241 KAAAEKALQEQAAQAAAAANNNTQATDASDQOAAAAADNTQAAQTGSDTOSAAQAVNN 300  
Db 241 KAAAEKALQEQAAQAAAAANNNTQATDASDQOAAAAADNTQAAQTGSDTOSAAQAVNN 300  
Qy 301 SDOESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 360  
Db 301 SDOESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 360  
Qy 361 ASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGVQGGQIQVQEAANYAGNQSIGNYRGW 420  
Db 361 ASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGVQGGQIQVQEAANYAGNQSIGNYRGW 420  
Qy 421 FNPGSVSVIYPN 432  
Db 420 FNPGSVSVIYPN 431

RESULT 7  
US-10-383-930-29  
; Sequence 29, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 29  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-29

Query Match 97.8%; Score 2070.5; DB 4; Length 431;  
Best Local Similarity 98.6%; Pred. No. 1.2e-118;  
Matches 426; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTTIG 60  
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTTIG 60

Qy 61 QVSALOTQOAELOAENORLEAQSATLGGQIOTLSSKI VARNESLKQOARSQAQKSNAAATSY 120  
Db 61 QVSALOTQOAELOAENORLEAQSATLGGQIOTLSSKI VARNESLKQOARSQAQKSNAAATSY 120  
Qy 121 INAIINSKSVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENOAINTVAANQE 180  
Db 121 INAIINSKSVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENOAINTVAANQE 180  
Qy 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAODOKATLVAQKAAAEAEARQAAAAQAAEA 240  
Db 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAODOKATLVAQKAAAEAEARQAAAAQAAEA 240  
Qy 241 KAAAEKALQEQAAQAAAAANNNTQATDASDQOAAAAADNTQAAQTGSDTOSAAQAVNN 300  
Db 241 KAAAEKALQEQAAQAAAAANNNTQATDASDQOAAAAADNTQAAQTGSDTOSAAQAVNN 300  
Qy 301 SDOESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 360  
Db 301 SDOESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 360  
Qy 361 ASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGVQGGQIQVQEAANYAGNQSIGNYRGW 420  
Db 361 ASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGVQGGQIQVQEAANYAGNQSIGNYRGW 420  
Qy 421 FNPGSVSVIYPN 432  
Db 420 FNPGSVSVIYPN 431

RESULT 8  
US-10-797-821-29  
; Sequence 29, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 29  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-821-29

Query Match 97.8%; Score 2070.5; DB 5; Length 431;  
Best Local Similarity 98.6%; Pred. No. 1.2e-118;  
Matches 426; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTTIG 60  
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTTIG 60  
Qy 61 QVSALOTQOAELOAENORLEAQSATLGGQIOTLSSKI VARNESLKQOARSQAQKSNAAATSY 120  
Db 61 QVSALOTQOAELOAENORLEAQSATLGGQIOTLSSKI VARNESLKQOARSQAQKSNAAATSY 120  
Qy 121 INAIINSKSVSDAINRVSAIREVVSANEKMLHQEQDQKAAVEQKHQENOAINTVAANQE 180

Db 121 INAIINSKSVDAINRVSAIREVWSANEKMLQQEQDKAAVEQKQENQAAINTVAANQE 180  
QY 181 TTAQNTNALNTQOAEAAQALNLQAELETTAQOKATLVAKKAAAEAAARQAAAAQAAAAEA 240  
Db 181 TTAQNTNALNTQOAEAAQALNLQAELETTAQOKATLVAKKAAAEAAARQAAAAQAAAAEA 240  
QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 300  
Db 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299  
QY 301 SDOESTTATAOPSPASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVNTWNGGGQWA 360  
Db 300 SDOESTTATAOPSPASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVNTWNGGGQWA 359  
QY 361 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVOGGQIQVQEAANYAGNOSIGNYRGW 420  
Db 360 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVOGGQIQVQEAANYAGNOSIGNYRGW 419  
QY 421 FNPGSVSYIYPN 432  
Db 420 FNPGSVSYIYPN 431

## RESULT 9

US-10-383-930-33  
; Sequence 33, Application US/10383930  
; Publication No. US20040127400A1

GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein

FILE REFERENCE: 25669-018  
CURRENT APPLICATION NUMBER: US/10/383,930

CURRENT FILING DATE: 2003-03-07

PRIOR APPLICATION NUMBER: 60/402,483

PRIOR FILING DATE: 2002-08-08

PRIOR APPLICATION NUMBER: 60/363,209

PRIOR FILING DATE: 2002-03-07

NUMBER OF SEQ ID NOS: 41

SOFTWARE: PatentIn version 3.2

SEQ ID NO 33

LENGTH: 431

TYPE: PRT

ORGANISM: Streptococcus mutans

US-10-383-930-33

Query Match 97.7%; Score 2066.5; DB 4; Length 431;  
Best Local Similarity 98.4%; Pred. No. 2e-118;  
Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAAQAVNTIQG 60  
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAAQAVNTIQG 60  
QY 61 QVSALQTOQAELEQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSAQKSNAAATSY 120  
Db 61 QVSALQTOQAELEQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSAQKSNAAATSY 120  
QY 121 INAIINSKSVDAINRVSAIREVWSANEKMLHQEQDKAAVEQKQENQAAINTVAANQE 180  
Db 121 INAIINSKSVDAINRVSAIREVWSANEKMLHQEQDKAAVEQKQENQAAINTVAANQE 180  
QY 181 TTAQNTNALNTQOAEAAQALNLQAELETTAQOKATLVAKKAAAEAAARQAAAAQAAAAEA 240  
Db 181 TTAQNTNALNTQOAEAAQALNLQAELETTAQOKATLVAKKAAAEAAARQAAAAQAAAAEA 240  
QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 300  
Db 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299  
QY 301 SDOESTTATAOPSPASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVNTWNGGGQWA 360

Db 300 SDOESTTATAOPSPASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVNTWNGGGQWA 359  
QY 361 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVOGGQIQVQEAANYAGNOSIGNYRGW 420  
Db 360 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVOGGQIQVQEAANYAGNOSIGNYRGW 419  
QY 421 FNPGSVSYIYPN 432  
Db 420 FNPGSVSYIYPN 431

## RESULT 10

US-10-797-821-33

; Sequence 33, Application US/10797821  
; Publication No. US20050031633A1

GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; APPLICANT: Taubman, Martin A.

; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens

FILE REFERENCE: 25669-020

CURRENT APPLICATION NUMBER: US/10/797,821

CURRENT FILING DATE: 2004-03-09

PRIOR APPLICATION NUMBER: 10/383,930

PRIOR FILING DATE: 2003-03-07

PRIOR APPLICATION NUMBER: 60/363,209

PRIOR FILING DATE: 2002-03-07

PRIOR APPLICATION NUMBER: 60/402,483

PRIOR FILING DATE: 2002-08-08

PRIOR APPLICATION NUMBER: 09/290,049

PRIOR FILING DATE: 1999-04-12

PRIOR APPLICATION NUMBER: 60/081,550

PRIOR FILING DATE: 1998-04-13

PRIOR APPLICATION NUMBER: 60/115,142

PRIOR FILING DATE: 1999-01-08

NUMBER OF SEQ ID NOS: 45

SOFTWARE: PatentIn version 3.2

SEQ ID NO 33

LENGTH: 431

TYPE: PRT

ORGANISM: Streptococcus mutans

US-10-797-821-33

Query Match 97.7%; Score 2066.5; DB 5; Length 431;  
Best Local Similarity 98.4%; Pred. No. 2e-118;  
Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAAQAVNTIQG 60  
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAAQAVNTIQG 60  
QY 61 QVSALQTOQAELEQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSAQKSNAAATSY 120  
Db 61 QVSALQTOQAELEQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSAQKSNAAATSY 120  
QY 121 INAIINSKSVDAINRVSAIREVWSANEKMLHQEQDKAAVEQKQENQAAINTVAANQE 180  
Db 121 INAIINSKSVDAINRVSAIREVWSANEKMLHQEQDKAAVEQKQENQAAINTVAANQE 180  
QY 181 TTAQNTNALNTQOAEAAQALNLQAELETTAQOKATLVAKKAAAEAAARQAAAAQAAAAEA 240  
Db 181 TTAQNTNALNTQOAEAAQALNLQAELETTAQOKATLVAKKAAAEAAARQAAAAQAAAAEA 240  
QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 300  
Db 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299  
QY 301 SDOESTTATAOPSPASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVNTWNGGGQWA 360  
Db 300 SDOESTTATAOPSPASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVNTWNGGGQWA 359  
QY 361 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVOGGQIQVQEAANYAGNOSIGNYRGW 420  
Db 360 ASAAAAGYRVGSTPSAGAVVWNDGGYGHVAVVTGVOGGQIQVQEAANYAGNOSIGNYRGW 419

```
; 421 FNPQSVSYIYPN 432
; |||||
Db 420 FNPQSVSYIYPN 431

RESULT 11
US-10-474-792-600
; Sequence 600, Application US/10474792
; Publication No. US20040236072A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Stephen
; APPLICANT: Zagursky, Robert
; APPLICANT: Nickbarg, Elliot
; APPLICANT: Winter, Lourie
; TITLE OF INVENTION: SURFACE PROTEINS OF STREPTOCOCCUS PYOGENES
; FILE REFERENCE: AM 100399
; CURRENT APPLICATION NUMBER: US/10/474,792
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 674
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 600
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-474-792-600

Query Match 51.1%; Score 1081; DB 5; Length 398;
Best Local Similarity 53.6%; Pred. No. 3.2e-58;
Matches 232; Conservative 56; Mismatches 109; Indels 36; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASODSKINNLTAQOQAAQAVNTTQ 60
DB 1 MKKRILSAVLVSGVTLGAATT---YGAEDLSKIAKQDSIIISNLTTEQAAQNVSAQ 57

QY 61 QVSALQTOQAELOAENORLEAQSATLGOQIOTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
DB 58 QVSSIQSEQDKLTARNTLEALSKEPQEIKALTSQIVARNEKLNQQAARSAYKNNETSGY 117

QY 121 INAINSKSVDAINRVSAIREVVSANEMHLHQEQDQKAAVEQKHQHOENQAINTVAANQE 180
DB 118 INALLNSKISIDVNVRLVAINRAVSANAKLLEQQKADKVSLEEQAAQNTAINTTAAANNA 177

QY 181 TIAQNTNALNTQOALEAQLNLOAELTTAQDOKATLVAQKAAAEAEARQAAQAAAE 240
DB 178 MABENQNTLRTOQNLVATANLALQLASATDEKANLVAQKEAEKAAAEALQAEQAAKV 237

QY 241 KAAAEAKALQEQAAQAAQAAANNNNTQATDASDQAAAAADNTQAAQTGDSSTDOQAAQAVNN 300
DB 238 KA-----QEQA-----QQAASVEAKSAITPAPQATPAAQSSNA 272

QY 301 SDQESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPVWVGNWNGGQWA 360
DB 273 IEPAAALTAPAA-PSAGPQT-----SYDSSNTYPVPGQCTGWGAKSLAPWAGNNGWNGGQWA 325

QY 361 ASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVVTGVO-GGQIQVOEANYAGNQSIGNYRG 419
DB 326 YSAQAAGYRTGTPMVGAIAVMNDGGYGHVAVVVEVQSASSIRVMESNYSGQYTDHRG 385

QY 420 WFNPGSVSYIYPN 432
DB 386 WFNPTGVTFIYPH 398

RESULT 12
US-10-472-928-4652
; Sequence 4652, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926W
```

```
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107659.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4652
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: secreted 45 kd protein (usp45)
; OTHER INFORMATION: Cellular location: outside
; OTHER INFORMATION: Similar to strain R6 sequence 15904062 (0.B+01)
US-10-472-928-4652

Query Match 49.8%; Score 1053; DB 5; Length 392;
Best Local Similarity 49.8%; Pred. No. 1.6e-56;
Matches 217; Conservative 74; Mismatches 93; Indels 52; Gaps 4;

QY 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASODSKINNLTAQOQAAQAVNTTQ 60
DB 1 MKKKILASLLSLSTVMVSVQVAVLTTHAETTDKIAAQDNKISNLTAAQOQAEAKQVDQIQE 60

QY 61 QVSALQTOQAELOAENORLEAQSATLGOQIOTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
DB 61 QVSAIQAEQSNLOAENDRLQAESKKEIGEITELSKNIVSRNQSLQKQAARSAGTNGAVTSY 120

QY 121 INAINSKSVDAINRVSAIREVVSANEMHLHQEQDQKAAVEQKHQHOENQAINTVAANQE 180
DB 121 INTIVNSKSITEAISRAVAMSEIVSANNKMLEQQKADKKAISEKQVANNDAINTVIANQQ 180

QY 181 TIAQNTNALNTQOALEAQLNLOAELTTAQDOKATLVAQKAAAEAEARQAAQAAAE 240
DB 181 KLADDAQALTTQOALEKAAELSLAAEKATASGEKASLLLEQKAAAEAEARAAVAEAAAYKE 240

QY 241 KAAAEAKALQEQAAQAAQAAANNNNTQATDASDQAAAAADNTQAAQTGDSSTDOQAAQAVNN 300
DB 241 KRASQQQSV-----LASANTNLTAQVOAVSESAAAPVRAKVRPT----- 279

QY 301 SDQESTTTAAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPVWVGNWNGGQWA 360
DB 280 -----YSTNASSYPITGECTGWGKTLAPWAGDYWGNGAQWA 314

QY 361 ASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVVTGVO-GGQIQVOEANYAGNQSIGNYRG 419
DB 315 TSAAAGPRTGSTPOVGAIAVMNDGGYGHVAVVTAVESTTRTIQVSESNYAGNRTIGNHRG 374

QY 420 WFNP-----GSVSYIY 430
DB 375 WFNPTTSEGFTYIY 390

RESULT 13
US-10-617-320-3230
; Sequence 3230, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
```

OPERATING SYSTEM: <Unknown>  
SOFTWARE: <Unknown>  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/617,320  
FILING DATE: 10-Jul-2003  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,433  
FILING DATE: 30-Jun-1998  
APPLICATION NUMBER: 60/085131  
FILING DATE: May 12, 1998  
APPLICATION NUMBER: 60/051553  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Arinello, Pamela Deneko  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 3230:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 399 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...399  
SEQUENCE DESCRIPTION: SEQ ID NO: 3230:  
US-10-617-320-3230

Query Match 49.8%; Score 1053; DB 5; Length 399;  
Best Local Similarity 49.8%; Pred. No. 1.7e-56;  
Matches 217; Conservative 74; Mismatches 93; Indels 52; Gaps 4;  
QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQVNTTQ 60  
DB 8 MKKILASLLSTVMVSVAVLTAAHTTDDKIAAQDNKISNLTAAQQAQVNDQIQE 67  
QY 61 QVSALQTOQAELOAENRLEAQSATLGOQIOTLSSKIVARNESLKQQAARSQAQSAATSY 120  
DB 68 QVSALQAEQSNLQAEQDRLEAQSATLGOQIOTLSSKIVARNESLKQQAARSQAQSAATSY 127  
QY 121 INALINSKVSDAINRVSAIREVVSANEMKHLHQEQDQKAAVEQKHQENQAQNTVAANQE 180  
DB 128 INTIVNSKSITEAIRVAAAMSVIVSANNKMLEQQKADKKATSEKQVANNDAINTVIANQQ 187  
QY 181 TTAQNTNALNTQQAQLEAAQLNLAELTTAQQQKATLVAAQKAAAEAEARQAAQAAAE 240  
DB 188 KLADDAQALTTQKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 247  
QY 241 KAAAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 300  
DB 248 KRAEQQSV-----LASANTNLTAQVQAVSESAAAPVRAKVRPT----- 286  
QY 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWA 360  
DB 287 -----YSTNASSYPIGECTWGVKSLAPWVGNWNGGOWA 321  
QY 361 ASAAAAGVVGSTPSAGAVVNDGGYGHVAVVTGVQV-GQIQVQEAANYAGNOSTGNV 419  
DB 322 TBAAGAFRTGTPQVGAICNDGGYGHVAVVTAVESTTRIQVESYVAGNRTIGNH 381  
QY 420 WFN-----GSVSYTY 430  
DB 382 WFNPTTTSBGFVTY 397

RESULT 14  
US-10-282-122A-57658

Sequence 57658, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 57658  
LENGTH: 524  
TYPE: PRT  
ORGANISM: Enterococcus faecium  
US-10-282-122A-57658

Query Match 27.7%; Score 585.5; DB 4; Length 524;  
Best Local Similarity 30.6%; Pred. No. 8.6e-28;  
Matches 161; Conservative 89; Mismatches 165; Indels 111; Gaps 14;  
QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQVNTTQ 60  
DB 1 MKKSLISAVMVCSTLTAVASPIAAADDFDQIQQDKIADLNKQADAQSDIALES 60  
QY 61 QVSALQTOQAELOAENRLEAQSATLGOQIOTLSSKIVARNESLKQQAARSQAQSAATSY 120  
DB 61 QVSEINTQADLLAKQDTRQESAQLVKDIADLQERIEKREDTIQKQAREQVNTSSNY 120  
QY 121 INALINSKVSDAINRVSAIREVVSANEMKHLHQEQDQKAAVEQKHQENQAQNTVAANQE 180  
DB 121 IDAVLNADSLADAIGRVQAMTTVMKANNNDLMEQQKQDKKAVEDKKAENDAKLKEAENQA 180  
QY 181 TTAQNTNALNTQQAQLEAAQLNLAELTTAQQQKATLVAAQKAAAEAEARQAAQAAAE 226  
DB 181 ALESQKGLLSKQADLNVLKTSLAQAQATAEKDKADLNKQAEAEQAEQAEQAEQAEQAE 240  
QY 227 AARQAAAAQAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAE 279  
DB 241 ARQAAQAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 297  
QY 280 NTQAAQTGDSTQSAQAQVNNSDQESTTA-----TAAQPSASSASTAAVAANTSS 329

```
Db      298 ATEESTPESSTEESTAPESSESTTESSATEESTTVPESSESTTESTTVPESSTTE 357
Qy      330 ANTPA-----GCTGWGKSLAP-----
Db      358 ESTTPAPTPTDQSVDTGNGT-GSSTPAPTPTPTPEQPKPVTAPAPSGSVNGAAIVAE 416
Qy      348 ---WGN--YWGNG-----GWAASAAAAGYRVG-STPSAGA 378
Db      417 AYKYGTPYVWVGKDPGSGDFGTFYVVMQVTGRDIGGWTVPQBSAGTKISVSOAKAGD 476
Qy      379 VAVW-NDGGYGHVAVYVVGQGOIOVQOEANYAGNOSIGNYRGWFNP 423
Db      477 LLFWSQGGTYHVAIALG--GGQ-YIHAPQPGESVKGVSQ-WFAP 518
```

```
RESULT 15
US-10-154-251-67
; Sequence 67, Application US/10154251
; Publication No. US20030092024A1
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Pritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 06286-060002
; CURRENT APPLICATION NUMBER: US/10/154,251
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 10/154,251
; PRIOR FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 67
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-154-251-67
```

```
Query Match      26.2%; Score 555; DB 4; Length 210;
Best Local Similarity 54.8%; Pred. No. 2.2e-26;
Matches 115; Conservative 44; Mismatches 51; Indels 0; Gaps 0;

Qy      1 MKKRILSAVLVSGVTLSSATTLISAVKADDPDAQLASQDSKINNLTAAQQAAQAVNTIQG 60
Db      1 MKKKILASLLSTVMVSQVAVLTTAAHETDDKIAAQDNKISNLTAAQQAAQAVNTIQG 60
Qy      61 QVSALQTOQAELEQAEORLEAQSATLIGQOIQTLSKIVARNESLKQARSQAQSNAAATSY 120
Db      61 QVSAIQAEQSNLQAEENDRLQAESKLEGEITELSKNIVSRNQSLEKQARSQAQNGAVTSY 120
Qy      121 INAIINKSVDAINRVSAIRVVSANEKMLHQEQODKAAVEQKHQENQAQAINTVAAQOE 180
Db      121 INTIVNSKSIETASIKRVAMSEIVSANNKMLEQKADKKAISEKQVANNDAINTVIANQQ 180
Qy      181 TIAQNTNALNTQQAQLAEALNQLAELTTA 210
Db      181 KLADDAQALTTKQAELEKAAELSLAAEKATS 210
```

Search completed: February 15, 2006, 18:34:53  
Job time : 93.071 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:30:27 ; Search time 8.33977 Seconds  
(without alignments)  
736.166 Million cell updates/sec

Title: US-10-797-821-32

Perfect score: 2116

Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNRYGMFNGSVIYIPN 432

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 107799 seqs, 14211699 residues

Total number of hits satisfying chosen parameters: 107799

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_New.\*

- 1: /cgn2\_6/prodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/prodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/prodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/prodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 5: /cgn2\_6/prodata/2/pubpaa/US05\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/prodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 7: /cgn2\_6/prodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/prodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID | Description        |
|------------|--------|-------------|--------|----|--------------------|
| 1          | 2070.5 | 97.8        | 431    | 7  | US-11-052-554A-210 |
| 2          | 1087   | 51.4        | 398    | 7  | US-11-052-554A-252 |
| 3          | 436.5  | 20.6        | 211    | 7  | US-11-052-554A-352 |
| 4          | 410    | 19.4        | 544    | 7  | US-11-052-554A-358 |
| 5          | 236    | 11.2        | 257    | 6  | US-10-793-626-3244 |
| 6          | 228    | 10.8        | 971    | 7  | US-11-052-554A-3   |
| 7          | 226.5  | 10.7        | 394    | 7  | US-11-052-554A-79  |
| 8          | 224    | 10.6        | 1236   | 6  | US-10-873-528-109  |
| 9          | 222.5  | 10.5        | 270    | 6  | US-10-485-517-413  |
| 10         | 214    | 10.1        | 149    | 6  | US-10-793-626-1682 |
| 11         | 214    | 10.1        | 157    | 6  | US-10-793-626-2870 |
| 12         | 193.5  | 9.1         | 2101   | 6  | US-10-857-780-23   |
| 13         | 188.5  | 8.9         | 1562   | 7  | US-11-052-554A-211 |
| 14         | 184.5  | 8.7         | 1095   | 6  | US-10-793-626-3154 |
| 15         | 183.5  | 8.7         | 758    | 6  | US-10-485-517-144  |
| 16         | 183.5  | 8.7         | 1448   | 6  | US-10-485-517-212  |
| 17         | 180.5  | 8.5         | 1122   | 6  | US-10-467-657-6112 |
| 18         | 180.5  | 8.5         | 1586   | 6  | US-10-821-234-901  |
| 19         | 178.5  | 8.4         | 1410   | 6  | US-10-878-556A-136 |
| 20         | 178    | 8.4         | 330    | 6  | US-10-485-517-415  |
| 21         | 177.5  | 8.4         | 1565   | 6  | US-10-467-657-2704 |
| 22         | 177    | 8.4         | 1107   | 6  | US-10-485-517-145  |
| 23         | 176    | 8.3         | 1290   | 6  | US-10-485-517-141  |
| 24         | 175    | 8.3         | 1126   | 6  | US-10-485-517-246  |
| 25         | 174    | 8.2         | 1634   | 7  | US-11-019-711-49   |

ALIGNMENTS

RESULT 1

US-11-052-554A-210  
; Sequence 210, Application US/11052554A  
; Publication No. US2005028866A1

GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07

; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 210

; TYPE: PRT

; LENGTH: 431

; ORGANISM: Streptococcus mutans UA159

US-11-052-554A-210

Query Match 97.8%; Score 2070.5; DB 7; Length 431;

Best Local Similarity 98.6%; Pred. No. 1.6e-113;

Matches 426; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Qy 1 MKKRILSAVLVSGVTLSATLSAVKADDDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60

Db 1 MKKRILSAVLVSGVTLSATLSAVKADDDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60

Qy 61 QVSALQTQQAELQAEQNRLQEQSQTIGQIQTLSSKIVARNESLKQQAQSAQSNATSY 120

Db 61 QVSALQTQQAELQAEQNRLQEQSQTIGQIQTLSSKIVARNESLKQQAQSAQSNATSY 120

Qy 121 INAINSKSVSDAIRVSAIREVVSANEKMLHQEQDQKAAVEQKQENQAANTVAANQE 180

Db 121 INAINSKSVSDAIRVSAIREVVSANEKMLHQEQDQKAAVEQKQENQAANTVAANQE 180

Qy 181 TTAQNTNALNTQQAQLQAEQLNQLQAEQLTTAQDQKATLVAKAAAEFAAQAQAQA 240

Db 181 TTAQNTNALNTQQAQLQAEQLNQLQAEQLTTAQDQKATLVAKAAAEFAAQAQAQA 240

Qy 241 KAAAEKALQEQAAQA 300

Db 241 KAAAEKALQEQAAQA 300

Qy 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPAGCQTGWGKSLAPWGVNNGGQWA 360

Db 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPAGCQTGWGKSLAPWGVNNGGQWA 360

Db 300 SDOESTTATEAPSSASTRAVAANTSSANTYPAGCCTGWKSLPVGNTWNGGQWA 359  
QY 361 ASAAAAGYRVGTPSAGAVVNDGGYGHVAVTVGGGQIQVQZANYAGNOSIGNYRGW 420  
Db 360 ASAAAAGYRVGTPSAGAVVNDGGYGHVAVTVGGGQIQVQZANYAGNOSIGNYRGW 419  
QY 421 FNPGSVSYIYPN 432  
Db 420 FNPGSVSYIYPN 431

RESULT 2  
US-11-052-554A-252  
; Sequence 252, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; PRIOR FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 252  
; LENGTH: 398  
; TYPE: PRT  
; ORGANISM: Streptococcus pyogenes MGAS8232  
US-11-052-554A-252

Query Match 51.4%; Score 1087; DB 7; Length 398;  
Best Local Similarity 53.8%; Pred. No. 1.5e-56;  
Matches 233; Conservative 56; Mismatches 108; Indels 36; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSATLSAVKADFDQAIASQDSKINNLTAAQOAAQAVNTTQ 60  
Db 1 MKKRILSAVLVSGVTILGAATT---VGAEDLSKIAKQDSIISNLATTEQKAAQNVSA 57

QY 61 QVSALOTQAELOAENORLEAQSATILGQIQITLSSKIVARNESLKQARSQAOKSNAATSY 120  
Db 58 QVSSLOQBQDKLTARNTLEALSKEFQEI KALTQI VARNESKLNQARSAYKNNETSGY 117

QY 121 INAIINSKVSDAINRVSAIREVWSANEKMLHQEQDKAAVQKHQEQNOAAINTVAA 180  
Db 118 INALLNSKISDVRLVAINRAVSNAKLLEQKADKVSLEKQAAQNTAINTIAANWA 177

QY 181 TIAQNTNALNTQOALEAAQLNQLABELTTPAQQOKATLVAKAAABEAAQAAAAQAA 240  
Db 178 MAEENQNTLRTQANLEAATANLALQASATEDKANLVAKAEAKAAEALAQOQAAKV 237

QY 241 KAAAEAKALOEAAQAAQAAANNNTQTADSDQOAAADNTQAAOTGSTDOSAAQAVN 300  
Db 238 KA-----QEQA-----QQAASVEAKSAITTPAQATPAAQSSNA 272

QY 301 SDOESTTATAOPSSASTRAVAANTSSANTYPAGCCTGWKSLPVGNTWNGGQWA 360  
Db 273 IEPALTAPAA-PSARPQT-----SYDSNTYPVGCTWGAKSLAPWAGNWNNGGQWA 325

QY 361 ASAAAAGYRVGTPSAGAVVNDGGYGHVAVTVGGQIQVQZANYAGNOSIGNYRG 419  
Db 326 YSAQAAGYRTGTPMVGAIVVNDGGYGHVAVVVEVQSASSIRVMESYSGRQYIADHRG 385

QY 420 FNPGSVSYIYPN 432  
Db 386 WFNPTGVTFIYPH 398

RESULT 3

US-11-052-554A-352  
; Sequence 352, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; PRIOR FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 352  
; LENGTH: 211  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans UA159  
US-11-052-554A-352

Query Match 20.6%; Score 436.5; DB 7; Length 211;  
Best Local Similarity 44.2%; Pred. No. 3.7e-19;  
Matches 95; Conservative 25; Mismatches 60; Indels 35; Gaps 7;

QY 247 KALQEAAQAAQAAANNNTQTADSDQOAAADNTQAAOTGDS-----TDQSAQA 297  
Db 3 KQFLEKAVFTVAAA-----TAATVVLGNKVADAD-TYTLQEGDSFFSVAQRVHMDAYELAS 56

QY 298 VNNSD-----QBSTTTATAAOPSASSA-----STAAVAANTSSANTYPAGCCTGWKVS 344  
Db 57 MNGKDIITSLILPGQTLTVNGSAAPNQAAAAPDTTQATTEINDANATYPVGCCTGWYKA 116

QY 345 LAPWVGNVWNGGQWAAASAAAGYRVGTPSAGAVVNDGGYGHVAVTVGV-QGGQIQV 403  
Db 117 VATWAGDWNGGQWASSASAGQYTVGNTPAVGSIMCMTDGGYGHVAVTVAVGEDGKVQV 176

QY 404 QEANYAGNQSIGNYRGWEN-----PGSVSYIYPN 432  
Db 177 LESNYKQOQWVDNVRGWFDPNNSGTPGSVSYIYPN 211

RESULT 4  
US-11-052-554A-358  
; Sequence 358, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; PRIOR FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 358  
; LENGTH: 544  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans UA159  
US-11-052-554A-358

Query Match 19.4%; Score 410; DB 7; Length 544;  
Best Local Similarity 25.3%; Pred. No. 3.8e-17;  
Matches 143; Conservative 78; Mismatches 186; Indels 158; Gaps 19;

QY 4 RILSAVLVSGVTLSATLSAVKADFDQAQIA-----SQDSKINNLTAAQOAAQAVNTTQ 59  
Db 2 RKLKVALFASILGLMAVSSYTAADTEDNQVTISHYNEQAGTFDYNVQAANGKTIQSID 61

|    |     |         |   |     |    |
|----|-----|---------|---|-----|----|
| Qy | 60  | QGVSA   | LOTQQAELQ-----AENORLEAQS-----                           | ATL | 86 |
|    |     | :       | :   | :   |    |
| Db | 62  | VAIWSE  | ENGQDDLKWYHASNDGNSQLTVHFAENHGSVKVGSYIAHAYITTDGNRVGNL    | 121 |    |
|    |     | :       | :   | :   |    |
| Qy | 87  | GOQLOT  | LSKKIVARNESLQO--QARSACKSNAATSYINAIINSKVSDAINRVS-----    | 139 |    |
|    |     | :       | :   | :   |    |
| Db | 122 | GKRKL   | SLS-----APQLSLKQGLQLFSLKPSAADQLFSAVWSDENGQDDLHWYTADAG   | 177 |    |
|    |     | :       | :   | :   |    |
| Qy | 140 | -----   | -----TREVVSANEKMLHQEQD-----KAAVE-QKHQENQAINTVAAN        | 178 |    |
|    |     | :       | :   | :   |    |
| Db | 178 | NTLAGY  | ANHKGYGTVHVHTYLKQNGKMPISAAQDIDIPKPKVKIQIDKINDTSYDVVVNN  | 237 |    |
|    |     | :       | :   | :   |    |
| Qy | 179 | -----   | -----QETIAQNTNALNTQQAOLFAAQLNLQAE--                     | 206 |    |
|    |     | :       | :   | :   |    |
| Db | 238 | VPPYISS | VAIPVWSEONGQDDLKWYQATKVADGIFKTTVYLKTRFELGNVQAHYGDSQ     | 297 |    |
|    |     | :       | :   | :   |    |
| Qy | 207 | -----   | -----LTTAQDOKATLVAQKAAAEAEAAQAAAAQAAAAEAKAAEAKAL        | 249 |    |
|    |     | :       | :   | :   |    |
| Db | 298 | LSKKLD  | GLGETHFNPVPSIINVEDPQVTI--DHYNINKGTFDVTVAE-----          | 348 |    |
|    |     | :       | :   | :   |    |
| Qy | 250 | QEQAAQ  | AAAAANNNT-----QATDASDQAAAAADNTQ--AAQFG-----             | 293 |    |
|    |     | :       | :   | :   |    |
| Db | 349 | QSTSA   | AVMSDANQANLYWEAKQLANGKAAITVDVQKHGNTGTSYVNVVYVHYNDGTTSG  | 408 |    |
|    |     | :       | :   | :   |    |
| Qy | 294 | AAQAVNS | DQESTTATAAQPSSASSNSTAAVAANTSANTYPAGQCTGWKSLAPVGNVW      | 353 |    |
|    |     | :       | :   | :   |    |
| Db | 409 | HVLA    | -----NOQLNQIVHYQPSA-----VRITAYMNEKNTPVPGQCTGWKELAPMIPNL | 459 |    |
|    |     | :       | :   | :   |    |
| Qy | 354 | GNGGOWA | ASAAAAAGYRVGSTPSAGAVAVWNDGGYGHVAYVTGVQ--GGQIQVOEANYAGNQ | 412 |    |
|    |     | :       | :   | :   |    |
| Db | 460 | GNGGOWA | STAVAKFGKIGTVPKVGNALCWSGGYGHVAYVTHVESNNRLOVKEANYKNQO    | 519 |    |
|    |     | :       | :   | :   |    |
| Qy | 413 | SIGNYR  | GWFNPD-----GSVSYIYPN                                    | 432 |    |
|    |     | :       | :   | :   |    |
| Db | 520 | YISNFR  | GWFDPTTSYGLRLLTIYPD                                     | 544 |    |
|    |     | :       | :   | :   |    |

```

RESULT 5
US-10-793-626-3244
; Sequence 3244, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 3244
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-3244

```

```

Qy 380 AVNWDGGYGHVAYVTGV-QGGQIQVQEQANYAGNCSIGNYRGWFPNGSVSYIY 430
Db 204 LOTSQAGAYGHVAYVEGVNSNGSIRVSEMMNYGHGAGVVTSTISASQAASNY 255

RESULT 6
US-11-052-554A-3
; Sequence 3, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3
; LENGTH: 971
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-3

```

```

RESULT 7
US-11-052-554A-79
; Sequence 79, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227

```

```

; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 79
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
; US-11-052-554A-79

Query Match      10.7%; Score 226.5; DB 7; Length 394;
Best Local Similarity 27.5%; Pred. No. 1.1e-06;
Matches 109; Conservative 61; Mismatches 178; Indels 49; Gaps 15;

QY      5  ILISAVLVSGVTLSSATTLSAVKADDFDAQIASQ-----DSKINNLTAAQQAAQAQVNTTIQ 59
      16  ILISAVL--HVILPAALINSS-----FDENIEASAGGGSSIDAVMVDSGAVVEQYKRMQ 68
QY      60  GOVSALOTQOAELOENQLEAQSATLGGQIQTLLSSKIVARNESIKQ-----QARSQAKS 114
      69  SOESS-----AKRSDEQRMKQEQAA-----EELREKQAAEQERLKQLEKERLAAQEQKK 118
QY      115  NAATSYINAIINSKVSDAINRVSAIREVVSANEKMLHQOQDQKAAVEQKHQENQAAINT 174
      119  QAEEAAKQAEKQAEAEAAKAAADAKAEADDKAAEEAAKKAADAKKAAEAAKA 178
QY      175  VVAHQETIAQNTVALNTQOAEALQNLQ---AELTTAQOKATILVAOKKAAAEAA--R 229
      179  AAEAQKAEAAAAALKKKAEAAEAAAEAAEARKKAAEAKAAADKKA---AEKAAAEKAAADK 235
QY      230  OAAQAQAAAEKAAAEKALQOAAQOAAQAAANNNNTQATDASDQOAAADNTQAAQTGDS 289
      236  KAAAEKAAADKKA--AKAAEAKAAAKAA-----EADDIFGELSSGKNAPKTTGG 286
QY      290  TDQSAQAQVNSDQBSTTATAAQ-PSASASTAAVAANTSSANTYPAGCQCTWGVKSLAP- 347
      287  AKGNNSPAGSGNTKNGASGADINNYAQIKSAIESKFYDASSYAGTKTCTLRK-LAPD 345
QY      348  -WVGNYWNGGQWA-ASAAAAGYRVGSTPSAGAVAV 382
      346  GMLLDIKPEGDPALCOAALAAAKLAKIKPPSQAVY 382

RESULT 8
US-10-873-528-109
; Sequence 109, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hanebro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21125WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 109
; LENGTH: 1236
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-109

Query Match      10.6%; Score 224; DB 6; Length 1236;
Best Local Similarity 23.3%; Pred. No. 6e-06;
Matches 91; Conservative 89; Mismatches 185; Indels 26; Gaps 10;

```

Db 187 STYSDAKYWASNAANDGYQVDNTPSVGAIMQSTPGPYGHVAYVERINGDGSILISEMNY 246

QY 409 AGNQSIGNYR 418  
|  
|  
|  
|

Db 247 ANGPYNMYR 256  
|  
|  
|  
|

RESULT 10  
US-10-793-626-1682  
; Sequence 1682, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1682  
; LENGTH: 149  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-1682

Query Match 10.1%; Score 214; DB 6; Length 149;  
Best Local Similarity 39.7%; Pred. No. 2e-06;  
Matches 50; Conservative 19; Mismatches 43; Indels 14; Gaps 5;

QY 287 GDSIQSAAQAVNNSDOESTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWG-KSL 345  
; Sequence 2870, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2870  
; LENGTH: 157  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-2870

Query Match 10.1%; Score 214; DB 6; Length 157;  
Best Local Similarity 39.7%; Pred. No. 2.1e-06;  
Matches 50; Conservative 19; Mismatches 43; Indels 14; Gaps 5;

QY 287 GDSIQSAAQAVNNSDOESTTATAQPSASSASTAAVAANTSSANTYPAGQCTGWG-KSL 345  
; Sequence 23, Application US/10857780  
; Publication No. US20050272043A1  
; GENERAL INFORMATION:  
; APPLICANT: ROTH, RICHARD B.  
; APPLICANT: BRAUN, ANDREAS  
; APPLICANT: KAMMERER, STEFAN M.  
; APPLICANT: NELSON, MATTHEW ROBERTS  
; APPLICANT: RENELAND, RIKARD HENRY  
; APPLICANT: HOYAL-WRIGHTSON, CAROLYN R.  
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS  
; FILE REFERENCE: SEQ-4069-CP  
; CURRENT APPLICATION NUMBER: US/10/857,780  
; CURRENT FILING DATE: 2004-05-28  
; PRIOR APPLICATION NUMBER: 10/723,681  
; PRIOR FILING DATE: 2003-11-25  
; PRIOR APPLICATION NUMBER: 60/490,234  
; PRIOR FILING DATE: 2003-07-24  
; PRIOR APPLICATION NUMBER: 60/525,239  
; PRIOR FILING DATE: 2003-11-25  
; NUMBER OF SEQ ID NOS: 4962  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 23  
; LENGTH: 2101  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-857-780-23

Query Match 9.1%; Score 193.5; DB 6; Length 2101;  
Best Local Similarity 24.1%; Pred. No. 0.00064;  
Matches 76; Conservative 58; Mismatches 113; Indels 69; Gaps 10;

QY 2 KKRILSAVLVSGVTLSATTLSA-----VKADDFDAQIASQDSKI-----NNLTAQQQAA 51  
; Sequence 2870, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2870  
; LENGTH: 157  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-2870

Query Match 10.1%; Score 214; DB 6; Length 157;  
Best Local Similarity 39.7%; Pred. No. 2.1e-06;  
Matches 50; Conservative 19; Mismatches 43; Indels 14; Gaps 5;

```
RESULT 13
US-11-052-554A-211
; Sequence 211, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 211
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-211

Query Match      8.9%; Score 188.5; DB 7; Length 1562;
Best Local Similarity 26.3%; Pred. No. 0.0089;
Matches 87; Conservative 57; Mismatches 116; Indels 71; Gaps 13;

QY 28 DDFDAQIASQDSKINNLTAQQAQAQVNTIQGVSAQTQQAELQAEQNRLEAQSAT-- 85
Db 151 DQYKSDVAHEAEVAKIKAKNQATKEQ-----YEKDMAHKAEEVERINAANAASK 200
QY 86 -----LGOQIOTLS--KIVARNESLQQAARSA-----OKSNAA--TSYINAIL-- 125
Db 201 TAYEAKLAQYQADLAARVOKTNAQAAQYKALAAQYAEKLRVQEAANAAKAAYDTAVAA 260
QY 126 NKSQSDAINRVSAIREVVSAEKMHLHQEQDQKAAVEQKHQENQAINTVAANQETIAQN 185
Db 261 NAKNT-----ETIAANEERIKRNATKABYETKLAQYQAEKRV-----QE 301
QY 186 TNALNTQQAQLEAQLNLTAQDQKATLVAQKAAAEAAQAAAAAQAQAAAEAKAAAE 245
Db 302 ANAAN-----EADYQAKLTAYQTELARVQKANADAKAAAYEAAVAANNAKNAALTAE 352
QY 246 AKALQEQAAQAA--ANNNTQATDASDQQAADNT--QAAQTGDSQDQAAQAVNNS 301
Db 353 NTAIKQRNENAKATYEAALKQYEAADLAAVKCANNAEADYQAKLTAYQTELARVOKAN-- 410
QY 302 DQESTTATAAQSASSASTAAVAANTSSANT 332
Db 411 ----ADAKAAVEAAVAAANNAALTAET-NT 436

RESULT 14
US-10-793-626-3154
; Sequence 3154, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3154
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
```

US-10-793-626-3154

```
Query Match      8.7%; Score 184.5; DB 6; Length 1095;
Best Local Similarity 22.5%; Pred. No. 0.001;
Matches 92; Conservative 60; Mismatches 166; Indels 91; Gaps 13;

QY 13 GVTLSATTLTSAVKADDFDAQIASQDSKINNLTAQQAQAQVNTIQGVSAQTQQAEL 72
Db 340 GMTTDTANNYSKKREAE-ETQKAAQIINNGDATEQQITETNRVQAQAINAKNDL 398
QY 73 QAEQNRLEAQSATLGGQIQTLSKIVA-RNESLQQAARSQAQKSAATSYINAIINS---- 127
Db 399 RADKSOLENAYNQLIQNVDTNGKKPASIQYQAAARQAIETQYNNAKSEAHQILENSPVS 458
QY 128 KVSVDAINRVSAIREVVSAEKMHLHQEQDQKAAVEQKHQENQA-----AINT 174
Db 459 NEVAQALQKVEAVQLKVNDIAHMLQNKNNNSALVTAQNQLQQAQVNDQPLTTGTTQDSINN 518
QY 175 VAA-----NOETIAQNTNA---LNTQQAQLEAAQNLNQ-----AELTTA 210
Db 519 YVAKRNEAQSAIRNAEAVINNGDATAKQISEKSVQALAHNDAKQOOLTADTTTELQTA 578
QY 211 QDQ-----KATLVAQKAAAEAAQAAAAAQAQAAAEAKAAAEAKALQEQAAQAA 259
Db 579 VQQLNRRGDTNNKKPRSNAYNAIKQSLQETQITSAKDNANAVIQPIRTVQE-VNNALQ 637
QY 260 ANNNTQATDASDQQAADN-----TQAQTGDSQDQ-- 293
Db 638 VNQLNQQLTEAINQLQPLSNNDALKAARLNLENKINQTVTDGMTQQSIEAYQNAKRVAQ 697
QY 294 -----AAQAVNNSD--QESTTATAAQSASSASTAAVAANTSSANTVP 334
Db 698 NESNTALALINNGDDEQIITETDRVQNQTNTLTQAI--NGLTVNKEP 744
```

RESULT 15

```
US-10-485-517-144
; Sequence 144, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynex Incorporated
; APPLICANT: Mond, James
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 144
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-144
```

```
Query Match      8.7%; Score 183.5; DB 6; Length 758;
Best Local Similarity 23.9%; Pred. No. 0.00076;
Matches 109; Conservative 60; Mismatches 143; Indels 145; Gaps 19;

QY 1 MKKRILSAVLVSGV-----TLSSA-----TTLSAVKADDFDA 32
Db 77 LKEQVQSATTLQGVQTVKNSSQTLTAMKGLRDSIANEATIKAGQNYTDASPNRNEYS 136
QY 33 QIASQDSKINNLTAQQAQAQVNTIQGVSAQTQQAELQAEQNRLEAQSATLGGQIQT 92
Db 137 AVTAAKAIIN-----QTSNPTMEPNTITQVTSQVTTKEQALNGARNLAQAKT-TAKNNLNN 191
QY 93 LSSKIVARNESL-----KQQAARSQAQKSAATSYIN 122
```



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 17:57:02 ; Search time 113.491 Seconds  
(without alignments)  
1668.609 Million cell updates/sec

Title: US-10-797-821-33

Perfect score: 2105

Sequence: 1 MKRILSAVLVSGVTLSAT.....SIGNVRGMFPGSVIYIYN 431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_21.\*

1: Geneseqp1980s.\*

2: Geneseqp1990s.\*

3: Geneseqp2000s.\*

4: Geneseqp2001s.\*

5: Geneseqp2002s.\*

6: Geneseqp2003as.\*

7: Geneseqp2003bs.\*

8: Geneseqp2004s.\*

9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description        |
|------------|--------|-------------|--------|-------|--------------------|
| 1          | 2105   | 100.0       | 431    | 7     | ADD93653 Streptoco |
| 2          | 2105   | 100.0       | 431    | 9     | ADx37276 Streptoco |
| 3          | 2096   | 99.6        | 431    | 7     | ADD93650 Streptoco |
| 4          | 2096   | 99.6        | 431    | 9     | ADx37273 Streptoco |
| 5          | 2092   | 99.4        | 431    | 7     | ADD93649 Streptoco |
| 6          | 2092   | 99.4        | 431    | 9     | ADx37272 Streptoco |
| 7          | 2092   | 99.4        | 431    | 9     | AEb91500 Microbial |
| 8          | 2075.5 | 98.6        | 432    | 7     | ADD93651 Streptoco |
| 9          | 2075.5 | 98.6        | 432    | 9     | ADx37274 Streptoco |
| 10         | 2066.5 | 98.2        | 432    | 7     | ADD93652 Streptoco |
| 11         | 2066.5 | 98.2        | 432    | 9     | ADx37275 Streptoco |
| 12         | 1238   | 58.8        | 447    | 5     | ABP29684 Streptoco |
| 13         | 1238   | 58.8        | 447    | 8     | ADU69524 S agalact |
| 14         | 1238   | 58.8        | 447    | 8     | ADV88392 Streptoco |
| 15         | 1238   | 58.8        | 447    | 8     | ADV81808 Streptoco |
| 16         | 1238   | 58.8        | 447    | 8     | ADv79645 Streptoco |
| 17         | 1092.5 | 51.9        | 398    | 9     | AEb91542 Microbial |
| 18         | 1085.5 | 51.6        | 398    | 5     | ABP25919 Streptoco |
| 19         | 1085.5 | 51.6        | 398    | 8     | ADP83884 S. pyogen |
| 20         | 1080   | 51.3        | 395    | 5     | ABP25918 Streptoco |
| 21         | 1057.5 | 50.2        | 392    | 6     | ABU02747 S. pneumo |
| 22         | 1057.5 | 50.2        | 392    | 8     | ADt50226 S_pneumon |
| 23         | 1057.5 | 50.2        | 392    | 8     | ADt50185 S_pneumon |
| 24         | 1057.5 | 50.2        | 399    | 8     | ADr94595 Novel S.  |

#### RESULT 1

ADD93653

ID ADD93653 standard; protein; 431 AA.

XX AC ADD93653;

XX DT 29-JAN-2004 (first entry)

XX DB Streptococcus mutans glucan binding protein-B.

XX KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.

XX OS Streptococcus mutans.

XX PN WO2003075845-A2.

XX PD 18-SEP-2003.

XX PF 07-MAR-2003; 2003WO-US006962.

XX PR 07-MAR-2002; 2002US-0363209P.

XX PR 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.

XX Smith DJ, Taubman MA;

XX WPI; 2003-845091/78.

XX GENBANK; AY046414.

XX Composition useful as vaccines for dental caries comprises a fragment of

XX a glucan binding protein-B binding to a major histocompatibility complex

XX class II protein.

XX Claim 5; Page 8-9; 49pp; English.

XX The present sequence is the protein sequence of the glucan binding

XX protein-B (GbpB) of Streptococcus mutans strain 5S93. The sequence

XX includes a number of human leucocyte antigen (HLA)-binding epitopes. The

XX invention provides immunogenic compositions and vaccines for dental

XX caries. The compositions comprise major histocompatibility complex (MHC)

XX class II protein-binding GbpB peptides, especially HLA-binding peptides,

XX covalently linked with peptide subunits of a glucosyltransferase. The

XX compositions are used in a claimed method of eliciting production of an

XX antibody in a mammal. Diepitopic or multi-epitopic polypeptides can be

XX prepared synthetically or by recombinant DNA technology. Antibodies

XX raised against MHC class II binding fragments of GbpB can be used in

#### ALIGNMENTS

|    |        |      |     |   |          |                    |
|----|--------|------|-----|---|----------|--------------------|
| 25 | 1057.5 | 50.2 | 399 | 9 | AEA58465 | Aea58465 Streptoco |
| 26 | 1056.5 | 50.2 | 392 | 8 | ADK47859 | Adk47859 Streptoco |
| 27 | 1056.5 | 50.2 | 392 | 8 | ADT50227 | Adt50227 S_pneumon |
| 28 | 671    | 31.9 | 461 | 2 | AAR14150 | Aar14150 MSP encod |
| 29 | 670    | 31.8 | 461 | 2 | AAR14530 | Aar14530 Usp45 pro |
| 30 | 661.5  | 31.4 | 456 | 5 | ABB55584 | Abb55584 Lactococc |
| 31 | 598    | 28.4 | 524 | 6 | ABU29734 | Abu29734 Protein e |
| 32 | 594    | 28.2 | 525 | 7 | ADC95468 | Adc95468 E. faeciu |
| 33 | 562    | 26.7 | 210 | 2 | AAy22579 | Aay22579 Bacterial |
| 34 | 526.5  | 25.0 | 482 | 9 | ADV16553 | Adv16553 E. faecal |
| 35 | 526.5  | 25.0 | 497 | 7 | ADH88105 | Adh88105 Enterococ |
| 36 | 507    | 24.1 | 449 | 2 | AAy00250 | Aay00250 Enterococ |
| 37 | 507    | 24.1 | 449 | 5 | ABP43469 | Abp43469 E faecali |
| 38 | 507    | 24.1 | 449 | 6 | ABU88497 | Abu88497 E. faecal |
| 39 | 507    | 24.1 | 449 | 6 | ABU13748 | Abu13748 Enterococ |
| 40 | 507    | 24.1 | 449 | 9 | ADV16734 | Adv16734 E. faecal |
| 41 | 507    | 24.1 | 449 | 9 | ADY39216 | Ady39216 Novel Ent |
| 42 | 507    | 24.1 | 450 | 7 | ADH87829 | Adh87829 Enterococ |
| 43 | 458    | 21.8 | 422 | 2 | AAy00251 | Aay00251 Enterococ |
| 44 | 458    | 21.8 | 422 | 5 | ABP43470 | Abp43470 E faecali |
| 45 | 458    | 21.8 | 422 | 6 | ABU88498 | Abu88498 E. faecal |

```

CC passive immunisation.
XX
SQ Sequence 431 AA;

Query Match      100.0%; Score 2105; DB 7; Length 431;
Best Local Similarity 100.0%; Pred. No. 7e-134;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQARSQAKSNAATSY 120
Db 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQARSQAKSNAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANKEMLQQEQDQKAAVEQKQENQAINTVAAQOE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANKEMLQQEQDQKAAVEQKQENQAINTVAAQOE 180
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAELETTAQQDKATLVAQKAAAEAEARQAAAAQAAEA 240
Db 181 TIAQNTNALNTQQAQLEAAQLNLQAELETTAQQDKATLVAQKAAAEAEARQAAAAQAAEA 240
QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNN 300
Db 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNN 300
QY 301 DOESTTATEAOPSASSASTAVVTANTSSANTYPAGCCTGWGKSLAPWGVNGNGGQWAA 360
Db 301 DOESTTATEAOPSASSASTAVVTANTSSANTYPAGCCTGWGKSLAPWGVNGNGGQWAA 360
QY 361 SAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYTVGGGQIQVQEANYAGNQSIGNYRGWF 420
Db 361 SAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYTVGGGQIQVQEANYAGNQSIGNYRGWF 420
QY 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431

RESULT 2
ADX37276
ID ADX37276 standard; protein; 431 AA.
XX
AC ADX37276;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus mutant glucan binding protein B variant #5.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus mutans.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
XX (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX
PI Smith DJ, Taubman MA;

WPI; 2005-151644/16.
XX
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
XX Claim 3; SEQ ID NO 33; 73pp; English.
XX
XX The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX
SQ Sequence 431 AA;

Query Match      100.0%; Score 2105; DB 9; Length 431;
Best Local Similarity 100.0%; Pred. No. 7e-134;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60
QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQARSQAKSNAATSY 120
Db 61 QVSALQTOQAELOAENQRLAQSATLGGQIQTLSSKIVARNESLKQARSQAKSNAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANKEMLQQEQDQKAAVEQKQENQAINTVAAQOE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANKEMLQQEQDQKAAVEQKQENQAINTVAAQOE 180
QY 181 TIAQNTNALNTQQAQLEAAQLNLQAELETTAQQDKATLVAQKAAAEAEARQAAAAQAAEA 240
Db 181 TIAQNTNALNTQQAQLEAAQLNLQAELETTAQQDKATLVAQKAAAEAEARQAAAAQAAEA 240
QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNN 300
Db 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNN 300
QY 301 DOESTTATEAOPSASSASTAVVTANTSSANTYPAGCCTGWGKSLAPWGVNGNGGQWAA 360
Db 301 DOESTTATEAOPSASSASTAVVTANTSSANTYPAGCCTGWGKSLAPWGVNGNGGQWAA 360
QY 361 SAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYTVGGGQIQVQEANYAGNQSIGNYRGWF 420
Db 361 SAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYTVGGGQIQVQEANYAGNQSIGNYRGWF 420
QY 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431

RESULT 3
ADX93650
ID ADD93650 standard; protein; 431 AA.
XX
AC ADD93650;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus mutans glucan binding protein-B.
XX
KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus mutans.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.

```

|    |  |  |
|----|--|--|
| XX | 07-MAR-2003; 2003WO-US006962.  |  |
| PF |  |  |
| XX | 07-MAR-2002; 2002US-0363209P.  |  |
| PR |  |  |
| XX | 08-AUG-2002; 2002US-0402483P.  |  |
| PR |  |  |
| XX | (FORS-) FORSYTH INST.  |  |
| XX |  |  |
| XX | Smith DJ, Taubman MA;  |  |
| PI |  |  |
| XX | WPI; 2003-845091/78.   |  |
| DR |  |  |
| XX | GENBANK; AY046411.   |  |
| DR |  |  |
| XX |  |  |
| PT | Composition useful as vaccines for dental caries comprises a fragment of   |  |
| PT | a glucan binding protein-B binding to a major histocompatibility complex   |  |
| PT | class II protein.  |  |
| XX |  |  |
| XX | Claim 5; Page 8; 49pp; English.  |  |
| PS |  |  |
| XX |  |  |
| CC | The present sequence is the protein sequence of the glucan binding         |  |
| CC | protein-B (GbpB) of Streptococcus mutans strain 3VP4. The sequence         |  |
| CC | includes a number of human leucocyte antigen (HLA)-binding epitopes. The   |  |
| CC | invention provides immunogenic compositions and vaccines for dental        |  |
| CC | caries. The compositions comprise major histocompatibility complex (MHC)   |  |
| CC | class II protein-binding GbpB peptides, especially HLA-B-binding peptides, |  |
| CC | covalently linked with peptide subunits of a glucosyltransferase. The      |  |
| CC | compositions are used in a claimed method of eliciting production of an    |  |
| CC | antibody in a mammal. Dieptopic or multiepitopic polypeptides can be       |  |
| CC | prepared synthetically or by recombinant DNA technology. Antibodies        |  |
| CC | prepared against MHC class II binding fragments of gbpB can be used in     |  |
| CC | passive immunisation.  |  |
| XX |  |  |
| XX | Sequence 431 AA;   |  |
| SQ |  |  |

| Query Match           | 99.6%;          | Score 2096;   | DB 7;     | Length 431; |
|-----------------------|-----------------|---|-----------|-------------|
| Best Local Similarity | 99.5%;          | Pred. No. 2.8e-133;   |           |             |
| Matches 429;          | Conservative 0; | Mismatches 2;   | Indels 0; | Gaps 0;     |
| Qy                    | 1               | MKKRILSAVLVSGVTLSSATLSAVKADDFDAQIASQDSKINNLTACQQAQAQVNTIQG  | 60        |             |
| Db                    | 1               | MKKRILSAVLVSGVTLSSATLSAVKADDFDAQIASQDSKINNLTACQQAQAQVNTIQG  | 60        |             |
| Qy                    | 61              | QVSLAQTCQAELOENORLEAQSATLGGQIOTLTSSKIVARNESLKQARSQKSNAAITSY | 120       |             |
| Db                    | 61              | QVSLAQTCQAELOENORLEAQSATLGGQIOTLTSSKIVARNESLKQARSQKSNAAITSY | 120       |             |
| Qy                    | 121             | INAIINSKVSDAINRVSATREVVSAANEKMLQQEQEDKAAVEQEQENQAAINTVAAEQE | 180       |             |
| Db                    | 121             | INAIINSKVSDAINRVSATREVVSAANEKMLQQEQEDKAAVEQEQENQAAINTVAAEQE | 180       |             |
| Qy                    | 181             | TIAQNTNALNTQQAELEAAQLNQAEITTAQDQKATLVAQKAAAEBAARQAAAAQAAEA  | 240       |             |
| Db                    | 181             | TIAQNTNALNTQQAELEAAQLNQAEITTAQDQKATLVAQKAAAEBAARQAAAAQAAEA  | 240       |             |
| Qy                    | 241             | KAAAEAKALQEAQAQAQAANNNNTQAATDASDQQAADNTQAAQTGDSQSAQAQVNNNS  | 300       |             |
| Db                    | 241             | KAAAEAKALQEAQAQAQAANNNNTQAATDASDQQAADNTQAAQTGDSQSAQAQVNNNS  | 300       |             |
| Qy                    | 301             | QDSETTATEAQPSASSASTAVVTANTSSANTYPAGQCTGWGKSLAPVWGNVTWNGGWAA | 360       |             |
| Db                    | 301             | QDSETTATEAQPSASSASTAVVAANTSANTYPAGQCTGWGKSLAPVWGNVTWNGGWAA  | 360       |             |
| Qy                    | 361             | SAAAGYRVGSTPSAGAVAVWMDGGYGHVAVVTGQGGQIQVEANTYAGNQSIGNYRGWF  | 420       |             |
| Db                    | 361             | SAAAGYRVGSTPSAGAVAVWMDGGYGHVAVVTGQGGQIQVEANTYAGNQSIGNYRGWF  | 420       |             |
| Qy                    | 421             | NPGSVSYIYPN   | 431       |             |
| Db                    | 421             | NPGSVSYIYPN   | 431       |             |

RESULT 4  
ADX37273

|    |  |
|----|--|
| ID | ADX37273 standard; protein; 431 AA.                                      |
| XX |  |
| AC | ADX37273;  |
| XX |  |
| DT | 21-APR-2005 (first entry)  |
| XX |  |
| DE | Streptococcus mutant glucan binding protein B variant #2.                |
| XX |  |
| KW | immunogenicity; immune stimulation; glucan binding protein-B;            |
| KW | microparticle; major histocompatibility complex; tooth disease.          |
| XX |  |
| OS | Streptococcus mutans.  |
| XX |  |
| FN | US2005031633-A1.   |
| XX |  |
| PD | 10-FEB-2005.   |
| XX |  |
| PF | 09-MAR-2004; 2004US-00797821.  |
| XX |  |
| PR | 13-APR-1998; 98US-0081550P.  |
| PR | 08-JAN-1999; 99US-0115142P.  |
| PR | 12-APR-1999; 99US-00290049.  |
| PR | 07-MAR-2002; 2002US-0363209P.  |
| PR | 08-AUG-2002; 2002US-0402483P.  |
| PR | 07-MAR-2003; 2003US-00383930.  |
| XX |  |
| PA | (SMIT// SMITH D J.   |
| PA | (TAUB/) TAUBMAN M A.   |
| XX |  |
| PI | Smith DJ, Taubman MA;  |
| XX |  |
| DR | WPI; 2005-151644/16.   |
| XX |  |
| PT | New composition comprising a fragment of a glucan binding protein-B      |
| PT | (GbpB) that binds to MHC class II protein, and a biocompatible           |
| PT | microparticle, useful for producing an antibody (claimed) for immunizing |
| PT | mammals against dental caries.   |
| XX |  |
| PS | Claim 3; SEQ ID NO 30; 73pp; English.                                    |
| XX |  |
| CC | The invention relates to a composition comprising a fragment of a glucan |
| CC | binding protein-B (GbpB) and a biocompatible microparticle, where the    |
| CC | fragment binds to a major histocompatibility complex (MHC), class II     |
| CC | protein. The composition is useful for producing an antibody for         |
| CC | immunizing mammals against dental caries. This sequence corresponds to a |
| CC | Streptococcus mutans GbpB protein of the invention.                      |

| AA | SQ  | Sequence              | 431                     | AA;                                     |
|----|-----|-----------------------|-------------------------|---|
|    |     | Query Match           | 99.6%;                  | Score 2096; DB 9; Length 431;           |
|    |     | Best Local Similarity | 99.5%;                  | Pred. No. 2.8e-133;                     |
|    |     | Matches 429;          | Conservative 0;         | Mismatches 2; Indels 0; Gaps 0;         |
| Qy | 1   | MKKRILSAVLVSGVTLS     | SATTL                   | SAVKADDPDAQIASQDSKINNLTACQQAQAQVNTIQ 60 |
| Db | 1   | MKKRILSAVLVSGVTLS     | SATTL                   | SAVKADDPDAQIASQDSKINNLTACQQAQAQVNTIQ 60 |
| Qy | 61  | QVSALOTQQAELQAE       | NORLEAQSATLGGQI         | OTTLSSKIVARNESIKQAQSAQKSNAAQTSY 120     |
| Db | 61  | QVSALOTQQAELQAE       | NORLEAQSATLGGQI         | OTTLSSKIVARNESLUKQAQSAQKSNAAQTSY 120    |
| Qy | 121 | INAIINSKVS            | DAINRVSAJREVVSANEKMLQQE | EQDQKAAVEQKQENQAINTVVAANQE 180          |
| Db | 121 | INAIINSKVS            | DAINRVSAJREVVSANEKMLQQE | EQDQKAAVEQKQENQAINTVVAANQE 180          |
| Qy | 181 | TTIAQNTNALNT          | QQAQLEAAQLNQAE          | LTTTACDQKATLVAQKAAAEAAARQAAAAQAAAAA 240 |
| Db | 181 | TTIAQNTNALNT          | QQAQLEAAQLNQAE          | LTTTACDQKATLVAQKAAAEAAARQAAAAQAAAAA 240 |
| Qy | 241 | KAAAEAKALQE           | QAAQAAAAANNNT           | QATDASDQQAADNTQAAQTGSTQGSAAQAQVNN 300   |
| Db | 241 | KAAAEAKALQE           | QAAQAAAAANNNT           | QATDASDQQAADNTQAAQTGSTQGSAAQAQVNN 300   |

QY 301 DQESTTATEAQSASSASTAVVTANTSSANTYPAGCTWGVKSLAPWVGNVWNGGQWAA 360  
Db |||||||||||||||||||  
QY 301 DQESTTATEAQSASSASTAAVAANTSSANTYPAGCTWGVKSLAPWVGNVWNGGQWAA 360  
Db |||||||||||||||||||  
QY 361 SAAAAAGYRVGTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVOEANYAGNQSIGNYRGWF 420  
Db |||||||||||||||||||  
QY 361 SAAAAAGYRVGTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVOEANYAGNQSIGNYRGWF 420  
QY 421 NPGSVSYIYPN 431  
Db |||||||||||||||||||  
QY 421 NPGSVSYIYPN 431

RESULT 5

ADD93649  
ID ADD93649 standard; protein; 431 AA.

AC ADD93649;

XX 29-JAN-2004 (first entry)

XX Streptococcus mutans glucan binding protein-B.

DE Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.

XX Streptococcus mutans.

| Key       | Location/Qualifiers          |
|-----------|------------------------------|
| FT Region | 6..25                        |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 16..35                       |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 33..52                       |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 37..56                       |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 48..67                       |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 52..71                       |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 88..107                      |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 113..132                     |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 117..136                     |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 137..156                     |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 174..193                     |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 194..213                     |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 214..233                     |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 248..267                     |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 289..308                     |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 306..325                     |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 311..330                     |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 349..368                     |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 365..384                     |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 383..402                     |
| FT Region | /note= "HLA-binding peptide" |
| FT Region | 403..422                     |
| FT Region | /note= "HLA-binding peptide" |

WO2003075845-A2.

XX

PD 18-SEP-2003.  
XX 07-MAR-2003; 2003WO-US006962.  
XX 07-MAR-2002; 2002US-0363209P.  
PR 08-AUG-2002; 2002US-0402483P.  
XX (FORS-) FORSYTH INST.  
XX Smith DJ, Taubman MA;  
PI WPI; 2003-845091/78.  
XX GENE BANK; AY046410.  
XX Composition useful as vaccines for dental caries comprises a fragment of  
PT a glucan binding protein-B binding to a major histocompatibility complex  
PT class II protein.  
XX  
PS Claim 6; Page 7; 49pp; English.  
XX  
CC The present sequence is the protein sequence of the glucan binding  
CC protein-B (GbpB) of Streptococcus mutans strain SK32. The sequence  
CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
CC invention provides immunogenic compositions and vaccines for dental  
CC caries. The compositions comprise major histocompatibility complex (MHC)  
CC class II protein-binding GbpB peptides, especially HLA-binding peptides,  
CC covalently linked with peptide subunits of a glucosyltransferase. The  
CC compositions are used in a claimed method of eliciting production of an  
CC antibody in a mammal. Dieptopic or multiptopic polypeptides can be  
CC prepared synthetically or by recombinant DNA technology. Antibodies  
CC raised against MHC class II binding fragments of GbpB can be used in  
CC passive immunisation.  
XX  
SQ Sequence 431 AA;

Query Match 99.4%; Score 2092; DB 7; Length 431;  
Best Local Similarity 99.3%; Pred. No. 5.3e-133;  
Matches 428; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTTQ 60  
Db |||||||||||||||||||  
QY 1 MKKRILSAVLVSGVTLSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTTQ 60  
Db |||||||||||||||||||  
QY 61 QVSALOTQOAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQARSQAQKSAATSY 120  
Db |||||||||||||||||||  
QY 61 QVSALOTQOAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQARSQAQKSAATSY 120  
Db |||||||||||||||||||  
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQEQDQKAAVEQKQEQENQAINTVAANQE 180  
Db |||||||||||||||||||  
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQEQDQKAAVEQKQEQENQAINTVAANQE 180  
Db |||||||||||||||||||  
QY 181 TIAQNTNALNTQOAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQARSQAQKSAATSY 240  
Db |||||||||||||||||||  
QY 181 TIAQNTNALNTQOAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQARSQAQKSAATSY 240  
Db |||||||||||||||||||  
QY 241 KAAAEKALQEQAAQAAQAAANNNTQATDASQOAAAADNTQAAQTGDSQSAQAQVNN 300  
Db |||||||||||||||||||  
QY 241 KAAAEKALQEQAAQAAQAAANNNTQATDASQOAAAADNTQAAQTGDSQSAQAQVNN 300  
Db |||||||||||||||||||  
QY 301 DQESTTATEAQSASSASTAVVTANTSSANTYPAGCTWGVKSLAPWVGNVWNGGQWAA 360  
Db |||||||||||||||||||  
QY 301 DQESTTATEAQSASSASTAVVTANTSSANTYPAGCTWGVKSLAPWVGNVWNGGQWAA 360  
Db |||||||||||||||||||  
QY 361 SAAAAAGYRVGTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVOEANYAGNQSIGNYRGWF 420  
Db |||||||||||||||||||  
QY 361 SAAAAAGYRVGTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVOEANYAGNQSIGNYRGWF 420  
Db |||||||||||||||||||  
QY 421 NPGSVSYIYPN 431  
Db |||||||||||||||||||  
QY 421 NPGSVSYIYPN 431

RESULT 6

```

ADX37272
ID  ADX37272 standard; protein; 431 AA.
AC  ADX37272;
XX
XX
XX  21-APR-2005 (first entry)
DE  Streptococcus mutant glucan binding protein B variant #1.
KW  immunogenicity; immune stimulation; glucan binding protein-B;
KW  microparticle; major histocompatibility complex; tooth disease.
XX
XX  Streptococcus mutans.
OS
XX
XX  US2005031633-A1.
XX
XX  10-FEB-2005.
XX
XX  09-MAR-2004; 2004US-00797821.
XX
XX  13-APR-1998; 98US-0081550P.
XX  08-JAN-1999; 99US-0115142P.
XX  12-APR-1999; 99US-00290049.
XX  07-MAR-2002; 2002US-0363209P.
XX  08-AUG-2002; 2002US-0402483P.
XX  07-MAR-2003; 2003US-00383930.
XX
XX  (SMIT/) SMITH D J.
XX  (TAUB/) TAUBMAN M A.
XX
XX  Smith DJ, Taubman MA;
XX
XX  WPI; 2005-151644/16.
XX
XX  New composition comprising a fragment of a glucan binding protein-B
XX  (GbpB) that binds to MHC class II protein, and a biocompatible
XX  microparticle, useful for producing an antibody (claimed) for immunizing
XX  mammals against dental caries.
XX
XX  Claim 3; SEQ ID NO 29; 73pp; English.
XX
XX  The invention relates to a composition comprising a fragment of a glucan
XX  binding protein-B (GbpB) and a biocompatible microparticle, where the
XX  fragment binds to a major histocompatibility complex (MHC) class II
XX  protein. The composition is useful for producing an antibody for
XX  immunizing mammals against dental caries. This sequence corresponds to a
XX  Streptococcus mutans GbpB protein of the invention.
XX
XX  Sequence 431 AA;

Query Match          99.4%; Score 2092; DB 9; Length 431;
Best Local Similarity 99.3%; Pred. No. 5.3e-133;
Matches 428; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
DB  1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
QY  61 QVSALQTQAELEQAEORLEAQSATLGGQIOTLSSKIVARNESLKQAEBSAOKSNAATSY 120
DB  61 QVSALQTQAELEQAEORLEAQSATLGGQIOTLSSKIVARNESLKQAEBSAOKSNAATSY 120
QY  121 INAINSKSVDAINRVSAIREVVSANERKMLQQEQDKAAVEQKQAEQENQAANTVAANO 180
DB  121 INAINSKSVDAINRVSAIREVVSANERKMLQQEQDKAAVEQKQAEQENQAANTVAANO 180
QY  181 TTAQNTNALNTQAQLEAQLNLQAELETTAQQKATLVAKAAAEFAAQAQAAQAAEA 240
DB  181 TTAQNTNALNTQAQLEAQLNLQAELETTAQQKATLVAKAAAEFAAQAQAAQAAEA 240
QY  241 KAAAEKALQEQAAQAAQAAANNNTQTADSDQAAADNTQAAQTGDSFEQSAQAQVNN 300
DB  241 KAAAEKALQEQAAQAAQAAANNNTQTADSDQAAADNTQAAQTGDSFEQSAQAQVNN 300

QY  301 DQESTTATEAQPSSASSASTAVVTANTSSANTYPAGQCTWGVKSLAPWVGNYMGNGQWAA 360
DB  301 DQESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNYMGNGQWAA 360
QY  361 SAAAGRYRGVSTPSAGAVAVNDGGYGHVAVYVTGVOGGQIQVQEAANYAGNQSIGNYRGWF 420
DB  361 SAAAGRYRGVSTPSAGAVAVNDGGYGHVAVYVTGVOGGQIQVQEAANYAGNQSIGNYRGWF 420
QY  421 NFGSVSYIYPN 431
DB  421 NFGSVSYIYPN 431

RESULT 7
AEB91500
ID  AEB91500 standard; protein; 431 AA.
XX
XX  AC  AEB91500;
XX
XX  20-OCT-2005 (first entry)
XX
XX  DE  Microbial pathogen adhesin protein sequence, SEQ ID NO:210.
XX
XX  KW  algorithm; adhesin; pharmaceutical; vaccine; drug screening;
XX  bordetella pertussis infection; antibacterial; pneumonia;
XX  antinflammatory; respiratory-gen.; gastric ulcer; antitumor;
XX  gastrointestinal-gen.; urinary tract infection; antimicrobial; uropathic.
XX
XX  OS  Streptococcus mutans.
XX
XX  PN  WO2005076010-A2.
XX
XX  PD  18-AUG-2005.
XX
XX  PF  07-FEB-2005; 2005WO-IN000037.
XX
XX  PR  06-FEB-2004; 2004IN-DE000173.
XX  20-JUL-2004; 2004US-0589227P.
XX
XX  PA  (COUL ) COUNCIL SCI & IND RES SOUTH AFRICA.
XX
XX  PI  Sachdeva G, Kumar K, Jain P, Brahmachari SK, Ramachandran S;
XX
XX  WPI; 2005-597835/61.
XX
XX  Computational method for identifying adhesin and adhesin like molecules,
XX  comprises computing sequence-based attributes of protein sequences using
XX  neural network software and training an artificial neural network.
XX
XX  Claim 16; SEQ ID NO 210; 402pp; English.
XX
XX  The present invention relates to a computational method (M1) for
XX  identifying adhesin and adhesin-like proteins, by computing the sequence-
XX  based attributes of protein sequences using five attribute modules of a
XX  neural network software, training an artificial neural network (ANN) for
XX  each of the computed five attributes, and identifying the adhesin and
XX  adhesin-like proteins having probability of being an adhesin (Pad) as
XX  equal or greater than 0.51. Also claimed is a set of 274 annotated genes
XX  encoding adhesin and adhesin-like proteins, having 274 fully defined
XX  11151 base pair (SEQ ID NO: 385-658) sequences; a set of 105 hypothetical
XX  genes encoding adhesin and adhesin-like proteins; a set of 279 annotated
XX  defined 306-15876 base pairs (SEQ ID NO: 659-763); a set of 279 annotated
XX  adhesin and adhesin-like proteins, having 279 fully defined 53-3716 base
XX  pair (SEQ ID NO: 1-279); a set of 105 hypothetical adhesin and adhesin-
XX  like proteins, having 105 fully defined 106-5291 base pair (SEQ ID NO:
XX  280-384) sequences; and a fully connected multilayer feed forward ANN (I)
XX  based on (M1). (M1) is useful for identifying adhesin and adhesin-like
XX  proteins, of therapeutic potential, and identifying and short-listing
XX  proteins for further testing in development of new vaccine formulations
XX  to eliminate diseases caused by various pathogenic organisms. (M1) is
XX  useful for identifying putative adhesins that are important in drug
XX  discovery and preventing therapeutics for whooping cough, pneumonia,

```

CC gastric ulcer and urinary tract infections. (M1) identifies adhesins from  
CC distantly related organisms, and from bacteria belonging to a wide  
CC phylogenetic spectrum. (M1) is capable of predicting adhesive nature of  
CC unique proteins. The present sequence is a microbial pathogen adhesin  
CC protein sequence.  
XX  
SQ Sequence 431 AA;

Query Match 99.4%; Score 2092; DB 9; Length 431;  
Best Local Similarity 99.3%; Pred. No. 5.3e-133;  
Matches 428; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60  
DB 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60  
QY 61 QVSALQTOQAEIQAENQRLAQSAATLGGQIQTLLSSKIVARNESLKQQAARSQAQSNAAATSY 120  
DB 61 QVSALQTOQAEIQAENQRLAQSAATLGGQIQTLLSSKIVARNESLKQQAARSQAQSNAAATSY 120  
QY 121 INAIINSKSVSDAINRVSAIREVWSANEKMLQQQEQDQKAAVEQKQENQAATNTVAANQE 180  
DB 121 INAIINSKSVSDAINRVSAIREVWSANEKMLQQQEQDQKAAVEQKQENQAATNTVAANQE 180  
QY 181 TTAQNTNALNTQQAQLEAAQNLQAEITTAODQKATLVAQKAAAEAAEQAAAQAQAAEA 240  
DB 181 TTAQNTNALNTQQAQLEAAQNLQAEITTAODQKATLVAQKAAAEAAEQAAAQAQAAEA 240  
QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDSDQQAADNTQAAQTGDSQSAQAQVNN 300  
DB 241 KAAAEAKALQEQAAQAAQAAANNNTQATDSDQQAADNTQAAQTGDSQSAQAQVNN 300  
QY 301 DOESTTATEAOPSASSASTAVTANTSSANTYPAGQCTGWGKSLAPWVGNVWNGGOWAA 360  
DB 301 DOESTTATEAOPSASSASTAVTANTSSANTYPAGQCTGWGKSLAPWVGNVWNGGOWAA 360  
QY 361 SAAAGYRVGSPSAGAVAVWNDGGYGHVAVYVTGQGGQIQVQEAANYAGNQSIGNYRGWF 420  
DB 361 SAAAGYRVGSPSAGAVAVWNDGGYGHVAVYVTGQGGQIQVQEAANYAGNQSIGNYRGWF 420  
QY 421 NPGSVSYIYPN 431  
DB 421 NPGSVSYIYPN 431

RESULT 8  
ADD93651  
ID ADD93651 standard; protein; 432 AA.  
XX AC ADD93651;  
XX 29-JAN-2004 (first entry)  
XX Streptococcus mutans glucan binding protein-B.  
XX Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.  
XX Streptococcus mutans.  
XX WO2003075845-A2.  
XX 18-SEP-2003.  
XX 07-MAR-2003; 2003WO-US006962.  
XX 07-MAR-2002; 2002US-0363209P.  
XX 08-AUG-2002; 2002US-0402483P.  
XX (FORS-) FORSYTH INST.  
XX Smith DJ, Taubman MA;  
XX WPI; 2003-845091/78.

DR GENBANK; AY046412.  
XX Composition useful as vaccines for dental caries comprises a fragment of  
PT a glucan binding protein-B binding to a major histocompatibility complex  
PT class II protein.  
XX  
PS Claim 5; Page 8; 49pp; English.

XX The present sequence is the protein sequence of the glucan binding  
CC protein-B (GbpB) of Streptococcus mutans strain 150P2. The sequence  
CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
CC invention provides immunogenic compositions and vaccines for dental  
CC caries. The compositions comprise major histocompatibility complex (MHC)  
CC class II protein-binding GbpB peptides, especially HLA-binding peptides,  
CC covalently linked with peptide subunits of a glucosyltransferase. The  
CC compositions are used in a claimed method of eliciting production of an  
CC antibody in a mammal. Pieptopic or multiptopic polypeptides can be  
CC prepared synthetically or by recombinant DNA technology. Antibodies  
CC raised against MHC class II binding fragments of GbpB can be used in  
CC passive immunisation.

XX Sequence 432 AA;

Query Match 98.6%; Score 2075.5; DB 7; Length 432;  
Best Local Similarity 98.6%; Pred. No. 6.9e-132;  
Matches 426; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60  
DB 1 MKKRILSAVLVSGVTLSATTLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTTIG 60  
QY 61 QVSALQTOQAEIQAENQRLAQSAATLGGQIQTLLSSKIVARNESLKQQAARSQAQSNAAATSY 120  
DB 61 QVSALQTOQAEIQAENQRLAQSAATLGGQIQTLLSSKIVARNESLKQQAARSQAQSNAAATSY 120  
QY 121 INAIINSKSVSDAINRVSAIREVWSANEKMLQQQEQDQKAAVEQKQENQAATNTVAANQE 180  
DB 121 INAIINSKSVSDAINRVSAIREVWSANEKMLQQQEQDQKAAVEQKQENQAATNTVAANQE 180  
QY 181 TTAQNTNALNTQQAQLEAAQNLQAEITTAODQKATLVAQKAAAEAAEQAAAQAQAAEA 240  
DB 181 TTAQNTNALNTQQAQLEAAQNLQAEITTAODQKATLVAQKAAAEAAEQAAAQAQAAEA 240  
QY 241 KAAAEAKALQEQAAQAAQAAA-NNNTQATDSDQQAADNTQAAQTGDSQSAQAQVNN 299  
DB 241 KAAAEAKALQEQAAQAAQAAAANNNTQATDSDQQAADNTQAAQTGDSQSAQAQVNN 300  
QY 300 SDQESTTATEAOPSASSASTAVTANTSSANTYPAGQCTGWGKSLAPWVGNVWNGGOWAA 359  
DB 301 SDQESTTATEAOPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNVWNGGOWAA 360  
QY 360 ASAAAAGYRVGSPSAGAVAVWNDGGYGHVAVYVTGQGGQIQVQEAANYAGNQSIGNYRGW 419  
DB 361 ASAAAAGYRVGSPSAGAVAVWNDGGYGHVAVYVTGQGGQIQVQEAANYAGNQSIGNYRGW 420  
QY 420 NPGSVSYIYPN 431  
DB 421 NPGSVSYIYPN 432

RESULT 9  
ADX37274  
ID ADX37274 standard; protein; 432 AA.  
XX AC ADX37274;  
XX 21-APR-2005 (first entry)  
XX Streptococcus mutant glucan binding protein B variant #3.  
XX immunogenicity; immune stimulation; glucan binding protein-B;  
KW microparticle; major histocompatibility complex; tooth disease.  
XX



QY 241 KAAAEKALQEAQAQAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299  
DB 241 KAAAEKALQEAQAQAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 300  
QY 300 SDOESTTATEAOPSASSASTAVVTANTSSANTYPAGQCTWGVKSLAPWVGNWGGQWA 359  
DB 301 SDOESTTATAAOPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWGGQWA 360  
QY 360 ASAAAAGYRVGSTPSAGAVAVWVNDGGYGHVAVVTGVGGQIQVQEAANYAGNQSIGNYRGW 419  
DB 361 ASAAAAGYRVGSTPSAGAVAVWVNDGGYGHVAVVTGVGGQIQVQEAANYAGNQSIGNYRGW 420  
QY 420 FNPGSVSIYYPN 431  
DB 421 FNPGSVSIYYPN 432

RESULT 11  
ADK37275  
ID ADK37275 standard; protein; 432 AA.  
XX  
AC ADK37275;  
XX  
DT 21-APR-2005 (first entry)  
XX  
DE Streptococcus mutant glucan binding protein B variant #4.  
XX  
KW immunogenicity; immune stimulation; glucan binding protein-B;  
KW microparticle; major histocompatibility complex; tooth disease.  
OS Streptococcus mutans.  
XX  
PN US2005031633-A1.  
XX  
PD 10-FEB-2005.  
XX  
PF 09-MAR-2004; 2004US-00797821.  
XX  
PR 13-APR-1998; 98US-0081550P.  
PR 08-JAN-1999; 99US-0115142P.  
PR 12-APR-1999; 99US-00290049.  
PR 07-MAR-2002; 2002US-0363209P.  
PR 08-AUG-2002; 2002US-0402483P.  
PR 07-MAR-2003; 2003US-00363930.  
XX  
PA (SMIT/) SMITH D J.  
PA (TAUB/) TAUBMAN M A.  
XX  
PI Smith DJ, Taubman MA;  
XX  
XX WPI; 2005-151644/16.  
XX  
XX New composition comprising a fragment of a glucan binding protein-B  
PT (GbpB) that binds to MHC class II protein, and a biocompatible  
PT microparticle, useful for producing an antibody (claimed) for immunizing  
PT mammals against dental caries.  
XX  
XX Claim 3; SEQ ID NO 32; 73pp; English.  
PS  
XX The invention relates to a composition comprising a fragment of a glucan  
CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
CC fragment binds to a major histocompatibility complex (MHC) class II  
CC protein. The composition is useful for producing an antibody for  
CC immunizing mammals against dental caries. This sequence corresponds to a  
CC Streptococcus mutans GbpB protein of the invention.  
XX  
SQ Sequence 432 AA;

Query Match 98.2%; Score 2066.5; DB 9; Length 432;  
Best Local Similarity 98.4%; Pred. No. 2.8e-131;  
Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;  
QY 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG 60

DB 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG 60  
QY 61 QVSALQTOQAELEAQNORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQKSNAAATSY 120  
DB 61 QVSALQTOQAELEAQNORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQKSNAAATSY 120  
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQQODKAAVEQKQENQAAINTVAANQE 180  
DB 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQQODKAAVEQKQENQAAINTVAANQE 180  
QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQDQKATLVAQKAAAEFAARQAAAAQAAAA 240  
DB 181 TTAQNTNALNTQQAQLEAAQLNLQAEELTTAQDQKATLVAQKAAAEFAARQAAAAQAAAA 240  
QY 241 KAAAEKALQEAQAQAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299  
DB 241 KAAAEKALQEAQAQAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 300  
QY 300 SDOESTTATEAOPSASSASTAVVTANTSSANTYPAGQCTWGVKSLAPWVGNWGGQWA 359  
DB 301 SDOESTTATAAOPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWGGQWA 360  
QY 360 ASAAAAGYRVGSTPSAGAVAVWVNDGGYGHVAVVTGVGGQIQVQEAANYAGNQSIGNYRGW 419  
DB 361 ASAAAAGYRVGSTPSAGAVAVWVNDGGYGHVAVVTGVGGQIQVQEAANYAGNQSIGNYRGW 420  
QY 420 FNPGSVSIYYPN 431  
DB 421 FNPGSVSIYYPN 432

RESULT 12  
ABP29684  
ID ABP29684 standard; protein; 447 AA.  
XX  
AC ABP29684;  
XX  
DT 02-JUL-2002 (first entry)  
XX  
DE Streptococcus polypeptide SEQ ID NO 8544.  
XX  
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
XX  
OS Streptococcus agalactiae.  
XX  
PN WO200234771-A2.  
XX  
PD 02-MAY-2002.  
XX  
PF 29-OCT-2001; 2001WO-GB004789.  
XX  
XX 27-OCT-2000; 2000GB-00026333.  
PR 24-NOV-2000; 2000GB-00028727.  
PR 07-MAR-2001; 2001GB-00005640.  
XX  
PA (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Telford J, Massignani V, Margarit Y Rosi, Grandi G, Fraser C;  
PI Tettelin H;  
XX  
DR WPI; 2002-352536/38.  
DR N-PSDB; ABN70315.  
XX  
XX New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.  
PS Claim 1; Page 3965; 4525pp; English.  
XX







Genomic nucleotide sequences encoding polypeptides of *Streptococcus agalactiae* for the development of vaccines, diagnostic tools, DNA chips and identification of therapeutic targets.

Claim 6; SEQ ID NO 2949; 439pp; French.

The present invention relates to novel *Streptococcus agalactiae* nucleotide sequences (I; ADV78860-ADV7898 and ADV83341-ADV85476) and novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The nucleotide sequences encode polypeptides of *S. agalactiae* involved in the synthesis of amino acids, cell membranes, intermediate (central) metabolism, energetic metabolism, fatty acid and phospholipid metabolism, nucleotide metabolism including purines, pyrimidines and/or nucleosides, regulatory functions, replication, transcription, translation, protein transport, adaptation to atypical conditions, sensitivity to medicines and/or analogues, functions related to transposons, biosynthesis of cofactors, prosthetic groups and transporters, cell membrane proteins and cellular machinery. (I) are useful for the detection and/or amplification of nucleic acids. Pharmaceutical composition comprising (I) or (II) are useful for treatment of a bacterial *S. agalactiae* infection. The complete genome of *Streptococcus agalactiae* is given in ADV81204. Note: The present patent is an equivalent for the basic patent FR262407A1, which contains only 2344 sequences.

Sequence 447 AA;

|                       |                  |  |            |             |
|-----------------------|------------------|--|------------|-------------|
| Query Match           | 58.8%;           | Score 1238;  | DB 8;      | Length 447; |
| Best Local Similarity | 60.7%;           | Pred. No. 2.7e-75;   |            |             |
| Matches 280;          | Conservative 43; | Mismatches 94;   | Indels 44; | Gaps 10;    |
|                       |                  |  |            |             |
| Qy                    | 1                | MKKRILSAVLVSGVTTLSAVKADDPDAQIASQDSKINNLTAAQQAAQAQVNTIQG        | 60         |             |
| Db                    | 1                | MKKRILSAVLVSGVTLGTAAV--TVNADDDFDSKIAATDSVINTLSGQAAAQNVTAIKG    | 58         |             |
|                       |                  |  |            |             |
| Qy                    | 61               | QVSGALQTQQAELQENORLEQAQSYATLGCQTQTTLSSKIVARNESLUKQAQSAQKNAATSY | 121        |             |
| Db                    | 59               | QVGALESQQSELEAQNQAQLEAVSQSLGQGETQLSLNKIVARNESLUKQVRSQAQGN-LTNY | 117        |             |
|                       |                  |  |            |             |
| Qy                    | 121              | INAIINSKSVSDAIRRVYSATREVVVSANEKMLQQEQODKAAVEOKQEQENQAINTVAAAOE | 180        |             |
| Db                    | 118              | INTILINSKSVSDAVNRVVAIREVVVSANEKMLAQEAEDKAALEAKQIENQAINTVAAKQ   | 177        |             |
|                       |                  |  |            |             |
| Qy                    | 181              | TIACNTNALMTQQAQLEAAQLNQIAELTTTQADQKATLVAQKAAAEAAARQAAAAQAAEA   | 240        |             |
| Db                    | 178              | AIENNKAAALATQRAQLEAAQLLESAQLTTVQNEKASLIQAKAQAEAAARKAAEAQAAEA   | 237        |             |
|                       |                  |  |            |             |
| Qy                    | 241              | KAAAEAKALQEAQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGDSTE--QSAQAQVN  | 298        |             |
| Db                    | 238              | KAQAEAKAQAEAVAKAQAAA-----QVESATAPITETVQTQPTETIKPSNLTA TS       | 286        |             |
|                       |                  |  |            |             |
| Qy                    | 299              | NSDQESTTATEA-----QPS-----ASSASTAVVTA-----NTSSANTYPAQG          | 336        |             |
| Db                    | 287              | SATTVATTTATATATNEPKVTQPSVVTKAVEAPKAVVSSTPRAVSKPVVRSYDSSNTYPMQG | 346        |             |
|                       |                  |  |            |             |
| Qy                    | 337              | CTWGVKSLAPWGNVYNGGOWASAAAAGVYGVGSTPSAGAVVW--NDGSGYGHVAVVTG     | 394        |             |
| Db                    | 347              | CTWAKMSAWGVNGVYNGVNAQWGSARAAGISVGTTPRVGAVAVWVPYDGGYIGHVAVVTS   | 406        |             |
|                       |                  |  |            |             |
| Qy                    | 395              | V-QGGQIQVQEAANYAGNOSIGNRWGNP---GSVSYIYPN                       | 431        |             |
| Db                    | 407              | VANNSSITOVNESNYAGNMSIGNRVSGSNPNSGSGVYIYPN                      | 447        |             |

Search completed: February 15, 2006, 18:03:01  
Job time : 114.491 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:03:21 ; Search time 20.9676 Seconds  
(without alignments)  
1977.789 Million cell updates/sec

Title: US-10-797-821-33

Perfect score: 2105  
Sequence: 1 MKKRILSAVLVSGVTLSSTAT.....SIGNRGMFNGSVSYIYPN 431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID       | Description        |
|------------|--------|-------------|--------|----------|--------------------|
| 1          | 1057.5 | 50.2        | 392    | 2 G95258 | secreted 45 kd pro |
| 2          | 1057.5 | 50.2        | 392    | 2 B98124 | general stress pro |
| 3          | 673    | 32.0        | 461    | 2 JN0097 | secreted 45K prote |
| 4          | 661.5  | 31.4        | 456    | 2 E8903  | hypothetical prote |
| 5          | 551    | 26.2        | 507    | 2 S05542 | hypothetical prote |
| 6          | 359.5  | 17.1        | 398    | 2 AC1763 | peptidoglycan lyti |
| 7          | 346    | 16.4        | 401    | 2 A11387 | peptidoglycan lyti |
| 8          | 319.5  | 15.2        | 473    | 2 F70031 | cell wall-binding  |
| 9          | 266    | 12.6        | 581    | 2 E75383 | conserved hypothet |
| 10         | 259    | 12.3        | 436    | 2 AH1387 | cell wall binding  |
| 11         | 257.5  | 12.2        | 461    | 2 H84099 | cell wall-binding  |
| 12         | 256.5  | 12.2        | 437    | 2 AB1763 | cell wall binding  |
| 13         | 238.5  | 11.3        | 528    | 2 B75310 | conserved hypothet |
| 14         | 230.5  | 11.0        | 421    | 2 JV0057 | tola protein - Esc |
| 15         | 227    | 10.8        | 971    | 2 B90835 | probable tail fibe |
| 16         | 227    | 10.8        | 973    | 2 C85693 | probable membrane  |
| 17         | 226    | 10.7        | 394    | 2 F90725 | membrane spanning  |
| 18         | 226    | 10.7        | 394    | 2 G85576 | membrane spanning  |
| 19         | 225.5  | 10.7        | 166    | 2 C90029 | hypothetical prote |
| 20         | 223.5  | 10.6        | 267    | 2 F90028 | hypothetical prote |
| 21         | 222.5  | 10.6        | 265    | 2 B98937 | hypothetical prote |
| 22         | 221.5  | 10.5        | 392    | 2 F96937 | cell wall-associat |
| 23         | 221.5  | 10.5        | 492    | 2 A28616 | M5 protein precurs |
| 24         | 220.5  | 10.5        | 4776   | 2 E95206 | cell wall surface  |
| 25         | 220    | 10.5        | 255    | 2 G90061 | hypothetical prote |
| 26         | 217.5  | 10.3        | 1122   | 2 G64887 | probable tail fibe |
| 27         | 215    | 10.2        | 688    | 2 A83179 | conserved hypothet |
| 28         | 214    | 10.2        | 536    | 2 D84325 | Htr17 transducer l |
| 29         | 212.5  | 10.1        | 483    | 2 A26297 | M6 protein - Strep |

RESULT 1

G95258  
secreted 45 kd protein [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: G95258  
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide  
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I.  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: G95258  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-392 <KUR>  
A:Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMY4; UNIPARC:UPI0000051B81; GB:AE005672; I  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP2216

|                       |       |  |           |            |    |        |     |
|-----------------------|-------|--|-----------|------------|----|--------|-----|
| Query Match           | 50.2% | Score  | 1057.5    | DB         | 2  | Length | 392 |
| Best Local Similarity | 50.6% | Pred.  | No. 2e-44 |            |    |        |     |
| Matches               | 220   | Conservative   | 73        | Mismatches | 91 | Indels | 51  |
| Gaps                  | 5     |  |           |            |    |        |     |
| Qy                    | 1     | MKKRILSAVLVSGVTLSSTATILSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG   | 60        |            |    |        |     |
| Db                    | 1     | MKKRILASLLSTVMVSVQAVLTTHAETDDKIAAQNKNISLTAQQAQAQVNTIQG         | 60        |            |    |        |     |
| Qy                    | 61    | QVSALQTOQAELQAENORLEAQSATLGOQIQTLSKIVARNESLKQARSQAQNAATSY      | 120       |            |    |        |     |
| Db                    | 61    | QVSATQAQSNLQAENDRLQAESKLEGEITLSKNIVSRNQSLEKQAARSATNGAVTSY      | 120       |            |    |        |     |
| Qy                    | 121   | INAIINSKSVDAINRVSAIREVVSANERKMLQQQFODKAAVFEQKQENQAQINTVAANQE   | 180       |            |    |        |     |
| Db                    | 121   | INTIVNSKITEAISRKAAMSIVSANNKMLEQKADKKAISEKQVANNDAINTVIANQQ      | 180       |            |    |        |     |
| Qy                    | 181   | TTAQNVTNALNTQAQLEAAQLNLQALHTTAQQAQKATLVAKAAAEAAQAAAAQAAAA      | 240       |            |    |        |     |
| Db                    | 181   | KLADDAQAQATTKQAELKAAELSLAAEKATAEKEKASLLEQKAAAEAAEAFAVAEAAAYK   | 239       |            |    |        |     |
| Qy                    | 241   | KAAAEKAKLQEAQAQAQAAAAANNNTQATDASDQAAAAADNTQAATQDSTEQSAQAQAVNNS | 300       |            |    |        |     |
| Db                    | 240   | -----EKRAQQQQSVLASANTNLTAQ-----                                | 267       |            |    |        |     |
| Qy                    | 301   | DOESTTATEAOPSASSASTAVVTANTSSANTYPAGQCTGWGKSLAPWPVGNWNGSQWAA    | 360       |            |    |        |     |
| Db                    | 268   | -----AAAPVRKVRPTYSTNASSYPIGECTGWGKTLAPWAGDNGNGAQWAT            | 315       |            |    |        |     |
| Qy                    | 361   | SAAAAGYRVGSTPSPAGAVAVNDGGYGHVAYVTGQG-CQIQVQEAANYAGNOSIGNYRGW   | 419       |            |    |        |     |

ALIGNMENTS

Db 316 SAAAGFRTGTPQVGAIACMNDGGYGHVAVVTAVESTTRIQVSESYNAGNRTIGNHRGW 375

QY 420 FNP-----GSVSYIY 429

Db 376 FNPFTTSEGFVYIY 390

RESULT 2

B98124

General stress protein GSP-781 [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004

C:Accession: B98124

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; E y, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.; A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; PMID:21429245; PMID:11544234

A:Accession: B98124

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-392 <KUR>

A:Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMY4; UNIPARC:UPI0000051B81; GB:AE007317; C:Genetics:

A:Gene: gsp-781

Query Match 50.2%; Score 1057.5; DB 2; Length 392;

Best Local Similarity 50.6%; Pred. No. 2e-44;

Matches 220; Conservative 73; Mismatches 91; Indels 51; Gaps 5;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTTIG 60

Db 1 MKKKILASLLSTVWVSVQVAVLTTHAETTTDDKIAAQDNKISNLTAQOQAEAKQVDDIOE 60

QY 61 QVSALQTOQAELOAENORLEAQSATILGQOIQTLSKIVARNESLKQOARSAQKSNAAATSY 120

Db 61 QVSAIQAEOSNQAENDRIQAESKKLEGEITELSKNIVSRNOSLEKQAQSAQTNGAVTSY 120

QY 121 INAIINSKSVSDAIRVSAIRVWSANEKMLQOQODKAAVPOKQOENQAAINTVAANQE 180

Db 121 INTIVNSKITEAISRVAAVMSIVSANNKMLSQKADKKAISEKVANNDAINTVIANQO 180

QY 181 TTAQNTNALNTQOAELEAAQLNLQAEELTTAQQOKATLVAKAAAEAEARQAAAAQAAEA 240

Db 181 KLADDAQALTTKQAEKLAELSLAAEKATPAEGEKASLLLEQKAAAEAEARAAVAEAAVK- 239

QY 241 KAAAEKALQEAQOAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAQAQVNN 300

Db 240 -----EKRAQQQSVLASANTNLTQA-----VQAVSES 267

QY 301 DOESTTATEAOPSASSASTAVVTANTSSANTYPAGQCTGWKSLAPWVGNVWNGGQWAA 360

Db 268 -----AAAPVRKVRPTVSTNASSYPIGECTVGTCTLPAPWAGDYWGNGAQWAT 315

QY 361 SAAAGYRVGSPSAGAVAVNDGGYGHVAVYTVGQG-QGIQVQEAANYAGNOSIGNYRGW 419

Db 316 SAAAGFRTGTPQVGAIACMNDGGYGHVAVVTAVESTTRIQVSESYNAGNRTIGNHRGW 375

QY 420 FNP-----GSVSYIY 429

Db 376 FNPFTTSEGFVYIY 390

RESULT 3

JN0097

secreted 45K protein precursor - Lactococcus lactis

C:Species: Lactococcus lactis

C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999

C:Accession: JN0097

R:van Asseldonk, M.; Rutten, G.; Oteman, M.; Siezen, R.J.; de Vos, W.M.; Simons, G. Gene 95, 155-160, 1990

Query Match 31.4%; Score 661.5; DB 2; Length 456;

Best Local Similarity 35.4%; Pred. No. 3e-25;

Matches 169; Conservative 103; Mismatches 138; Indels 67; Gaps 15;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTTIG 60

Db 1 MKKKILASLLSTVWVSVQVAVLTTHAETTTDDKIAAQDNKISNLTAQOQAEAKQVDDIOE 60

QY 61 QVSALQTOQAELOAENORLEAQSATILGQOIQTLSKIVARNESLKQOARSAQKSNAAATSY 120

A:Title: Cloning of usp45, a gene encoding a secreted protein from Lactococcus lactis sul

A:Reference number: JN0097; PMID:91071599; PMID:2123812

A:Accession: JN0097

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-461 <VAN>

A:Cross-references: UNIPARC:UPI000016D72B; GB:M35374

Query Match 32.0%; Score 673; DB 2; Length 461;

Best Local Similarity 35.2%; Pred. No. 8.6e-26;

Matches 169; Conservative 104; Mismatches 139; Indels 68; Gaps 13;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTTIG 60

Db 1 MKKKILASLLSTVWVSVQVAVLTTHAETTTDDKIAAQDNKISNLTAQOQAEAKQVDDIOE 59

QY 61 QVSALQTOQAELOAENORLEAQSATILGQOIQTLSKIVARNESLKQOARSAQKSNAAATSY 120

Db 60 KVDLSLQKQKTSKQIAKIESEAKALNAQIATLINESIKERTKTLEAQAARSQVNSSATNY 119

QY 121 INAIINSKSVSDAIRVSAIRVWSANEKMLQOQODKAAVPOKQOENQAAINTVAANQE 180

Db 120 MDVAVNSKSLTDVIOKVTAIATVSSANKQMLEQKEKELSKSETVKCKNYNQFVLSQ 179

QY 181 TTAQNTNALNTQOAELEAAQLNLQAEELTTAQQOKATLVAKAAAEAEARQAAAAQAAEA 240

Db 180 SLDQOAEELTQOAELEAAQLNLQAEELTTAQQOKATLVAKAAAEAEARQAAAAQAAEA 239

QY 241 KAAAEKALQEAQOAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQ-----SAAQ 295

Db 240 Q-----QKEAAQAASTAATAKAVEAATSSASASSSQAPQVSTSTDTNTTSNASN 291

QY 296 AVNNSDQSTTATEAOPSASSASTAVVTANTSS-----ANTYP 333

Db 292 SSNSSSSSSSSSSSSSSSSSSSSSSAGNTNSGTSTGNTGTTGGSGINSPIGNPYA 351

QY 334 AGOCT---WG-----VKSLAPWVGNVWNGGQWAAASAAAG--YRVGSPSAGAVAV 380

Db 352 GGGCTDYWQVFAAQGIYIRIMP-----GNGGQWASNGPAQGVLVHVGAAP--GVVAS 403

QY 381 -----WPDGGYGHVAVYTVGQG-QGIQVQEAANYAGNOSIGNYRMFPGSVSYIYPN 431

Db 404 SPSADFEVGYANSPYGHVAIVKSVNSDGTITIKEGGY-GTTWGHGR-TVASGAVTFLMPN 461

RESULT 4

E86903

hypothetical protein usp45 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis

C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 09-Jul-2004

C:Accession: E86903

R:Boletini, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlic

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssi

A:Reference number: A86625; PMID:21235186; PMID:11337471

A:Accession: E86903

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-456 <STO>

A:Cross-references: UNIPROT:Q9CDJ1; UNIPARC:UPI00000C6BFC; GB:AE005176; PID:g12725296; P

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: usp45

Query Match 31.4%; Score 661.5; DB 2; Length 456;

Best Local Similarity 35.4%; Pred. No. 3e-25;

Matches 169; Conservative 103; Mismatches 138; Indels 67; Gaps 15;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTTIG 60

Db 1 MKKKILASLLSTVWVSVQVAVLTTHAETTTDDKIAAQDNKISNLTAQOQAEAKQVDDIOE 59

QY 61 QVSALQTOQAELOAENORLEAQSATILGQOIQTLSKIVARNESLKQOARSAQKSNAAATSY 120

```
Db 60 KVDLSQKQASTKAQIAKTESEAKALNAQIATLINESIAERTKTLEAQASQVNSSATNY 119
Qy 121 INAINSKSVSDAINRVSAIREVVSANERKMLQOEODKAABVEQKQOENQAAINTVAANOE 180
Db 120 MDRAVNSKSLTDVIOKVTHAIATVSSANKQWLFQOEKEQELQKSEKQVTKVNTQNFVLSQ 179
Qy 181 TTAQNTNALNTQQAQLEAAQLNQLBELTTAQDQKATLVAKAAAEAEARQAAAAQAAAEA 240
Db 180 SLDSSQAQELTSQAELKVATLNTQATIAQDKQSLDLDEKAAAEKAAQEAQAKQAAYEA 239
Qy 241 KAAAEKALQEAQAQAQAANNTQATDASDQQAADNTQAAQTDGSTEQ-SAAQAVNN 299
Db 240 Q-----QKEAAKAQAATAAT--TKAVEEATSTVSSQASQSSSSNTSSNTSSN 289
Qy 300 SQEESTTATEAOPSASSAST-----AVVTANTSS-----ANTYPACQ 336
Db 290 SSSSSSSSSSSSSSSSSSSNGGSGTNTGNNNAAGTNTGGSSSGSINSTPIANPYAGG 349
Qy 337 CT---WG-----VKSLAPVWVNGVNGGQWAAASAAAG--YRVGSTPSSAGAVAV--- 380
Db 350 CTDYVWQYPAAGQIYIRNIMP-----GNGGQWATNGPAQGVHLVVGGAAP--GVIASP 401
Qy 381 -----WNDGGYGHVAVYTCV-QGGQIQVOEANYAGNQSIGNYRGWFPGSVSVIYPN 431
Db 402 ADFVGYANSYPGHVAIVKSVNSNGTITIKEGGY-GTTWGHGR-TVSAGVTFLLMPN 456

RESULT 5
S05542
Hypothetical protein, 54K - Enterococcus faecium
C:Species: Enterococcus faecium
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
C:Accession: S05542
R:Fuerst, P.; Moesch, H.U.; Solioz, M.
Nucleic Acids Res. 17, 6724, 1989
A:Title: A protein of unusual composition from Enterococcus faecium.
A:Reference number: S05542; MUID:89385998; PMID:2780297
A:Accession: S05542
A:Molecule type: DNA
A:Residues: 1-507 <FUE>
A:Cross-references: UNIPROT:P13692; UNIPARC:UPI000016F6RC; GB:X16421; EMBL:M26048; NID:5
A>Note: The authors translated the codon CGT for residues 221 and 223 as Lys

Query Match 26.2%; Score 551; DB 2; Length 507;
Best Local Similarity 28.9%; Pred. No. 7.3e-20;
Matches 147; Conservative 93; Mismatches 166; Indels 102; Gaps 13;

Qy 10 LVSGVTLSSATLSAVKADDFPAQIASQDSKINNLTAAQQAQAVNTTQGVSAALQTQQ 69
Db 1 MLSSIALTAGVSPIAAAADDFDSIQQQDKKIADLQNAQSAQSQIEALEGQVSAINTKA 60
Qy 70 ABLQENQRLAQSATLGGQIQTLSSKIVARNESLKQAARSQAQSKNAATSYINAIINSK 129
Db 61 QDLTKQDTLRKESAQLKQEIKDQBRIEKREATIQKQARETOVQVNTSSNYIDAVLNADS 120
Qy 130 VSDAINRVSAIREVVSANERKMLQOEODKAABVEQKQOENQAAINTVAANQETIAQNTAL 189
Db 121 LADAVGRIQAMSTIVKANQDLVQQQKEDQAEVKAENAKQELADNQAALQESQKGLD 180
Qy 190 NTQQAQLEAAQLNQLBELTTAQDQKATLVAKA-----AAEAARQAAAAQ 235
Db 181 LAKQADNLVLTSLAAEQATBEDKKADLNRKKAEEAEQARIEQARLAEQARQAAQAEK 240
Qy 236 AAEEAKAAAEKALQEQAAQAQAANNTQATDASDQQAADN-----TQ 281
Db 241 AEKBAEQAAQAQAAQTAQALSSASTTTTSSSAQSSSEESKAPESSTTESTSTSTTTE 300
Qy 282 AAQTDGSTEQSA--QAVNNSDQESTTATEAQPSSASTAV-----VT 323
Db 301 NSTGSSSTESSSTESTVPESTQESTPANTESSSSSNTVNNNTNNSTNNSTNTT 360
Qy 324 ANTSSANTY-----PAGQCTWGVKSLA-----PWGNYWGNG----- 355
```

```
Db 361 NNNNNNTVTPAPTPTPTPPAPAPNPSPGSGVNGAAIVAEAYKYIGTPYV---WGGKDP 417
Qy 356 -----GOWAASAAAAGYRVG-STPSAGAVAVVND--GGYGHVAVYVTG 394
Db 418 FDCSGFTRVYLVQVTRDIGGTVTPQESAGTKLSVQAKAGDLLFWGSAGTYHVAISLG 477
Qy 395 VGGGQIQVOEANYAGNQSIGNYRGWFPN 422
Db 478 --GGQ-YIHAPQPGENVKVGVSQ-WYTP 501

RESULT 6
AC1763
peptidoglycan lytic protein P45 [imported] - Listeria innocua (strain Clp11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AC1763
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Fsihi, H.;
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Maqueno, E.; Maitournam, A.; Mat
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AC1763
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-398 <GLA>
A:Cross-references: UNIPROT:Q927Y8; UNIPARC:UPI00000CC935; GB:AL592022; PIDN:CAC97875.1;
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: spl

Query Match 17.1%; Score 359.5; DB 2; Length 398;
Best Local Similarity 27.0%; Pred. No. 9.7e-11;
Matches 113; Conservative 89; Mismatches 146; Indels 71; Gaps 16;

Qy 1 MKKRILSAV-LVSGVTLSSATLSAVKADDFPAQIASQDSKINNLTAAQQAQAVNTTQ 59
Db 1 MKNTFIALSLAAVLSLTPAFTNVPA--DVNTDIGNQDKKINDIKSKTKGLQSDLSLV 58
Qy 60 GOVSALQTQQAELQAEQRLAQSATLGGQIQTLSSKIVARNESLKQAARSQAQSKNAATS 119
Db 59 ADLEKAQEKAKSLQGEFDQTKELQNLQDIKINERIKERETVLKERAMQKTSNSNA 118
Qy 120 YINAIINSKSVSDAINRVSAIREVVSANERKMLQOEOD-----KAAVEQKQOENQAAI 172
Db 119 YLEVLDAENLSDLVGRVSAVNLVSDSKSILEDDQKDEKALKTKQTAVKKQEEQATAI 178
Qy 173 NTVAAANQETIAQNTNALNTQQAQLEAAQLNQLBELTTAQDQKATLVAKAAAEAEARQAA 232
Db 179 HEFRAQQ-----NKIEAQAKAEKAIVAQLAADQASAEKAGLVSE--DKAAKE-- 226
Qy 233 AAQAAAEKAAAEKALQEQAAQAQAANNTQATDASDQQAADNNTQAAQOTGDSSTES 292
Db 227 -----ATARATALRE-----ATDANVGQOTT--NTNA--SSNSKTSN 259
Qy 293 AAQAVNNSDQESTTATEAOPSASSASTAVVTANTSSANTYPAG-----QCTWGVKSL- 344
Db 260 KYESTNNSEAPSA----ATPSSGGYSAMITAAARALGKPYLSIGATGPSAFDCS-GFTSYA 314
Qy 345 --APWGVNNGVNGGQWAAASAAAAGYRVGSTPSAGAVAVVN-DGGYGHVAVYVTGQGGQI 400
Db 315 FRAAGVSLPRISGGYVAAASKIS----ASQAEPGLVFNFGGIAHVIYVG--GGQM 367
```

```
RESULT 7
A11387
peptidoglycan lytic protein P45 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
```

C:Accession: A11387  
R:Glaeser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Fsihi, H.; Science 294, 849-852, 2001  
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.; Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: A11387  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-401 <GLA>  
A:Cross-references: UNIPROT:Q9RE04; UNIPARC:UPI00000D019D; GB:NC\_003210; PIDN:CAD00583.1  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: spl

Query Match 16.4%; Score 346; DB 2; Length 401;  
Best Local Similarity 26.3%; Pred. No. 4.4e-10;  
Matches 110; Conservative 93; Mismatches 148; Indels 68; Gaps 15;

QY 1 MKKRILSAVLVSGVTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 59  
DB 1 MKKNTFIAISLAAVISLTPAFTTNVFA--DVNTDIQNQDKINDIKSKKTDLQSDLSGLV 58

QY 60 GOVSALOTQQAELQENQLEAQSATLGGQIQTLSKIVARNESLKQARSQAQKNAATS 119  
DB 59 ADLEKAEQAKSLOGEFDKTKELKLNEDIKSINERIKERTVLKERARAMQKTSNSNA 118

QY 120 YNAILNSKVSQDAINRVSAIREVVSANEKMLQQOEOBKAAVEQKQ-----QENQA-AI 172  
DB 119 YLEVLDIAENLDLVRVSAVNVQLVDSKSLLEDQONDEKALKTKQTAVKKQEQEQAAT 178

QY 173 NTVAAANQETIAQNTNALNTQQAELAAQLNLAELTTAQQOKATLVAQKAAAEAAQAA 232  
DB 179 HEYEAQQ-----NKIEAQKAEKAIVAQLASDQASAEAKAGLVSR--DKAAKE-- 226

QY 233 AAQAAAEKAAAEKALQEAQAAQAAANNNTQATDASDQQAQAAADNTQAAQTGSTQSS 292  
DB 227 -----ATARATALRE-----ATSSNVGQSSSDSTSPSKSKSTTKN 263

QY 293 AAQAVNNSDOESTTATEAQPSSASSASTAVVTANTSSANTYPAG-----OCTGWGKSL- 344  
DB 264 VASNDNNSAPSAT-----PSSGGVSAMISAANAQLGRPYSLGAGPSAFDCS-GFTSYA 317

QY 345 --APWGVNYWNGGQMAASAAAAGYRVGSTPSSAGAVAVVNDG-GYGHVAVYTVGGGQI 400  
DB 318 FRAAGVSLPRTSGGGYAAASKIS---ASQAKPGDLVFFNYGGGIAHVGIYVG--GGQM 370

RESULT 8  
F70031  
cell wall-binding protein homolog yvce - *Bacillus subtilis*  
C:Species: *Bacillus subtilis*  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C:Accession: F70031  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter, C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler, ich, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F. Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapides, A.; Lardinois, A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel Y, M.; Ogawa, K.; Ogiwara, C.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpersira, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K A:Authors: Yoshikawa, H.F.; Zumstede, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: F70031  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-473 <KUN>  
A:Cross-references: UNIPROT:P40767; UNIPARC:UPI0000060AE1; GB:Z99121; GB:AL009126; NID:g  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yvce

Query Match 15.2%; Score 319.5; DB 2; Length 473;  
Best Local Similarity 24.0%; Pred. No. 9.9e-09;  
Matches 111; Conservative 95; Mismatches 186; Indels 71; Gaps 11;

QY 1 MKKRILSAVLVSGVTLS-----ATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQ 54  
DB 1 MKKSLITLGLASVIGTSSFLIPFTSKTASAEITLKKOKIESKQSEV---ASSIEAKEKE 57

QY 55 VNTIQGVSAALQTOQAEL-----QAENQRLQAQSATLGGQIQTLSKIVAR 100  
DB 58 LTELQENQSKIEKELKDINDKALDTSNKIEDKKEENDTKBEIKKLKKEIKETEAREIEKR 117

QY 101 NESLKQAARSQAQKNAATSYNIAIINSKVSQDAINRVSAIREVVSANEKMLQQOQSDKAA 160  
DB 118 NEILKKRVSLQESGSGQYIDVLGSGTSFGDFISRATAVSSIVDADKDLIKQEQDKAK 177

QY 161 VE-----QKQENQAAINTVAANQETIAQNTNALNTQQAELAAQLNLAELTTAQQO 213  
DB 178 LEDSEADLNDLKKVQAALAKLETWKQDLKQLEKQKLFDEAKASQKTKAKAISLSE 237

QY 214 KATLVAQKAAAEKALQEAQAAAEKALQEAQAAQAAANNNTQATDASDQ 273  
DB 238 ASELANQKANT-----AEQARIKKEQEAALIKKQEEAQKA-----SDETQTDSDQ 285

QY 274 AAAADNTQAAQTGSTEOSAAQAVNNSQESTTATEAQPSSASSASTAVVTANTSSANTYP 333  
DB 286 TATTESSKASSDDSDNSNNGSSNGSSNGSSKSSGSGNSNGTIVTSNNGGIEG 345

QY 334 AGQCTGWGKSLAPVGVNYWNGG-----OWA-ASAAAAGYRVGSTTSPA 375  
DB 346 AIVSGSSIVGGSP-----YKFGGRTQSDINNRIFDCCSFVRWAYASAGVNLGPVGGTTTD 401

QY 376 GAVAVVNDGGYGHVAVYTVGGGQIQQVQEAANYAGNQSIGNYRG 418  
DB 402 TLV-----GRQAVSASEMKRGDL-VFFDITYKTNGHVGIYLG 437

RESULT 9  
E75383  
conserved hypothetical protein - *Deinococcus radiodurans* (strain R1)  
C:Species: *Deinococcus radiodurans*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: E75383  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; F, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M. Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: E75383  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-581 <WHI>  
A:Cross-references: UNIPROT:Q9RU45; UNIPARC:UPI00000C195A; GB:AE001998; GB:AE000513; NID: A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR1549  
A:Map position: 1

Query Match 12.6%; Score 266; DB 2; Length 581;  
Best Local Similarity 23.5%; Pred. No. 4.7e-06;  
Matches 122; Conservative 75; Mismatches 173; Indels 150; Gaps 17;

QY 16 LSSATTL SAVKADDFDAQIASQ-----DSKINNLTAAQQAAQAQVNTIQGQ 61



```

Db      40 LAGGDTLSTRLEQLQRIEQRRLSVQKQKRLQTEVRARLRLQLNAQRAALRDLTALTS 99
QY      62 VSALOTQQAELQENORLEQAQATLQ---QIOTLSSKIVARNESLKQQAARSQKS---N 115
Db      100 VTDLENELADVLA--RVTAATRALRETEAQIRVTSQV---EALKVDARAVMKALYRA 152
QY      116 AATSYINAIINSGSVSDAINRVSAIREVVSANEKMLQ---QEQDKAAV 161
Db      153 RNTQYMRLLSQNSISDMLIRLDYANNAGQRNVNMEQLRGAALTTQQLRQRQSDAL 212
QY      162 EOKQENQ-----AAINTVAANQETIAQNTNALNTQQAQ----- 195
Db      213 RGLQGEQQTALBLRDRTRQDALAELQSAQGOQAVAVRTOQAQALTAQITDSLGNV 272
QY      196 -----LEAAQLNLQAE-----LTTAQDOKATLVA-- 219
Db      273 VERTFLERERRRLEERREAEARRIREAQERAKAEARLARIRAEQERKAREAAEA 332
QY      220 --QKAAAEARQAAAAQAAAEAKAAAEAKALQE-----QAAQAAANNNNTQATDSDQ 273
Db      333 ARQKALAE--ARAAQAKVAEAKAKAEAKAEAKAEAKAEAKAEAKAEAKAEAKAEAKAE 390
QY      274 AAAANTQAAQTGDSQSAQAQVNNSDQESTTA-----TEAQPAS-- 315
Db      391 AAARAQAQQAQAAAAAEARAEARAAQAAQQAQQAQKVRQAVAREQDILQTOQQAQOE 450
QY      316 ---SASTAVVTANTSSANTYPAGQGTGWKSLAPWGNVGNWGQWAAASAAAGYRVGST 372
Db      451 KQLAELAPLPAISRDLGFPLP---GKVAAP-----YGTGAQWEVLVSAGSGRAVAA 502
QY      373 PSAGAVA-----VWMDGGYGHVAVYVTVGGGQIQV 402
Db      503 QTGNVIASAYAAALGWILLDHGNSVITGYFGLQDTLVEV 542

RESULT 10
AH1387
cell wall binding proteins homolog lmo2504 [imported] - Listeria monocytogenes (strain E
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AH1387
R:Glauber, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlant,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AH1387
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-436 <GLA>
A:Cross-references: UNIPROT:Q8Y4E2; UNIPARC:UPI0000055221; GB:NC_003210; PIDN:CAD00582.1
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2504

Query Match      12.3%; Score 259; DB 2; Length 436;
Best Local Similarity 22.8%; Pred. No. 7.5e-06;
Matches 97; Conservative 90; Mismatches 152; Indels 86; Gaps 14;

QY      6 LSAVLVSG-VTLSSATLTSVAV-----KADDFDAQIASODSKNNLTAAQQAQAV 55
Db      11 LSLIIISAPLTGAHADTINDMKRQNEIEQKSEIDKNDKNSLNLHLESAEKDAKEL 70
QY      56 NTIQGV-----SALQTQQAELQENORLEQAQATLQ---QIOTLSSKIVARNESLKQQAARS 111
Db      71 ESLMKSLDPTNKKLKEQEDKVSSNEKLA---KLQKEMEKLNRDIRDQRKVLNRAI 126
QY      112 QKSNATSYINAIINSGSVSDAINRVSAIREVVSANEKMLQOQ---EQDKAAVEQKQENQ 169

```

```

Db      127 QTTGTATSYLDMIFRADDFKELVDRVTVVVSALVKADQNTMQDKQDDKLKVAESTSEKK 186
QY      170 AA-INTVAANQETIAQNTNALNTQQAELQAAQLNLQAELETTAQDQKATLVAQKAAAEAA 228
Db      187 LENLKVLAVELEVSXNNMESQKQKNDLVMLAN-KKDLT--KSEOTLLASEQCALTDEE 243
QY      229 ROAAAAQAAAAAEAKAAAEAKALQEQAQAQAAANNTQATDASDQAAAAADNTQAAQTGDS 288
Db      244 KLASNIAGEKAKQEAIAKAEKRMQEAANA-----S 276
QY      289 TQSAQAQAVNNSDQSTTATEAQPASSA-----STAVVTANTSSANTYPAGQGTGW 341
Db      277 SAKSAHV---KQPSSSSNEATETVSSGGGQFIRKPSGILTSGFSERTN----- 322
QY      342 KSLAPWGNVWGN-----GGQWAAASAAAGYRVGSTPSAGAVAVNNDDGGYGHVAVYTG 394
Db      323 ----PVTGKYESHKGQDIAGGTVTVSAAASGTVVFSFGFASGSGF---GGYGVVVKIDH 375
QY      395 VGGQ 399
Db      376 GNGFQ 380

RESULT 11
H84099
cell wall-binding protein BH3600 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: H84099
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, F.; Hiran
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: AB3650; MUID:20512582; PMID:11058132
A:Accession: H84099
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-461 <STO>
A:Cross-references: UNIPROT:Q9KX4; UNIPARC:UPI00000C424C; GB:AP001519; GB:BA0000004; NID
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3600

Query Match      12.2%; Score 257.5; DB 2; Length 461;
Best Local Similarity 22.1%; Pred. No. 9.5e-06;
Matches 98; Conservative 75; Mismatches 180; Indels 91; Gaps 11;

QY      31 DAQIASQSKNNLTAAQQAQAAQAVNTIQGVSAQTQQAELQENORLEQAQATLQOI 90
Db      56 EAELEKEVEKELGDTTAEIERLDKEVETSGKIQEKREIEEVQAEIEELK-----EQI 108
QY      91 QTLSSKIVARNESLKQQAARSQAQSNAAATSYINAIINSGSVSDAINRVSAIREVVSANEKM 150
Db      109 EILEERIAERDELLKDRARAWYQNGSGSIDYLEVILGAKSFGDFLDRVSAHSVIA----- 162
QY      151 LQQQODKAAVQKQEQENQAALNTVAANQETIAQNTNALNTQQAELQAAQLNLQAELETTA 210
Db      163 ----EQDRGILEAHIEDHRL-----EAKAQVEEKLETLEGLHLEVE-NLMAQLEEQ 210
QY      211 QDOKATLVAQKAAAEARQAAAAQAAAEAKAAAEAKALQEQ-----AAQAQAA 259
Db      211 QKEKEKVMGELASRDELHGLDLENDLELRQKEKALQEYELWKKQEEERKAAEKAA 270
QY      260 ANNTQATDASDQAAAAADNTQAAQTGDSQSAQAQVNNSDQESTTATEAQPAS-SAS 318
Db      271 AEAAQAQQAQSSGGSSGSSGSSGSSGTTSRNSGSSGGGGG-----TGSVPSSSGSGF 326
QY      319 TAVVTANTSSANTYPAGQGTGWKSLAPWGNVGNWGQWAAASAAAGYRVGSTPSAGAV 378
Db      327 MRPATGDISSPFGYRTHPVT-QQRKLHAGIDIRRGNRNVVVAAYDGTWVQST----- 379
QY      379 AVWMDGGYGHVAV-----TGVOGG-- 398
Db      380 --YSSGGYGNVWIIAHSYNGRQVTLIAHLETRSVSAGORVSKGTQIGIMTGLTGPH 437

```





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 17:57:21 ; Search time 123.476 Seconds  
(without alignments)  
2462.693 Million cell updates/sec

Title: US-10-797-821-33  
Perfect score: 2105  
Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNRYGRWNPQSVSIYPN 431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID           | Description        |
|------------|--------|-------------|--------|-----------------|--------------------|
| 1          | 2105   | 100.0       | 431    | 2 Q938V0_STRMU  | Q938V0 streptococc |
| 2          | 2096   | 99.6        | 431    | 2 Q9AG98_STRMU  | Q9AG98 streptococc |
| 3          | 2092   | 99.4        | 431    | 2 Q938V3_STRMU  | Q938V3 streptococc |
| 4          | 2092   | 99.4        | 431    | 2 Q8DMW3_STRMU  | Q8DMW3 streptococc |
| 5          | 2075.5 | 98.6        | 432    | 2 Q938V2_STRMU  | Q938V2 streptococc |
| 6          | 2066.5 | 98.2        | 432    | 2 Q938V1_STRMU  | Q938V1 streptococc |
| 7          | 1238   | 58.8        | 447    | 2 Q9AKA4_STRAG  | Q9AKA4 streptococc |
| 8          | 1238   | 58.8        | 447    | 2 Q8E2H1_STRAS  | Q8E2H1 streptococc |
| 9          | 1238   | 58.8        | 447    | 2 Q8E7X9_STRAS  | Q8E7X9 streptococc |
| 10         | 1095   | 52.0        | 474    | 2 Q5M6K4_STRTP2 | Q5M6K4 streptococc |
| 11         | 1092.5 | 51.9        | 398    | 2 Q5XEL1_STRP6  | Q5XEL1 streptococc |
| 12         | 1092.5 | 51.9        | 398    | 2 Q7CNQ7_STRP8  | Q7CNQ7 streptococc |
| 13         | 1092.5 | 51.9        | 398    | 2 Q8P318_STRP3  | Q8P318 streptococc |
| 14         | 1087.5 | 51.7        | 485    | 2 Q5M212_STRTP1 | Q5M212 streptococc |
| 15         | 1085.5 | 51.6        | 398    | 2 Q8A1Z8_STRPY  | Q8A1Z8 streptococc |
| 16         | 1057.5 | 50.2        | 392    | 2 Q8DMY4_STRR6  | Q8DMY4 streptococc |
| 17         | 1057.5 | 50.2        | 392    | 2 Q97N55_STRPN  | Q97N55 streptococc |
| 18         | 947    | 45.0        | 211    | 2 QZAS7_STRMU   | QZAS7 streptococc  |
| 19         | 666    | 31.6        | 461    | 1 USP45_IACLIC  | P22865 lactococcc  |
| 20         | 661.5  | 31.4        | 456    | 2 Q9CDJ1_LACLA  | Q9CDJ1 lactococcc  |
| 21         | 598    | 28.4        | 524    | 2 Q9K2J9_ENTFC  | Q9K2J9 enterococcu |
| 22         | 583    | 27.7        | 516    | 1 P54_ENTFC     | P13692 enterococcu |
| 23         | 566.5  | 26.9        | 576    | 2 Q9KJ3_ENTHR   | Q9KJ3 enterococcu  |
| 24         | 549.5  | 26.1        | 482    | 2 Q5M5M6_STRTP2 | Q5M5M6 streptococc |
| 25         | 548.5  | 26.1        | 470    | 2 Q8QE3_9LACT   | Q8QE3 lactococcc   |
| 26         | 529.5  | 25.2        | 482    | 2 Q93LK4_ENTFA  | Q93LK4 enterococcu |
| 27         | 514    | 24.4        | 461    | 2 Q5GSA7_STRTR  | Q5GSA7 streptococc |
| 28         | 511    | 24.3        | 449    | 2 Q93LK3_ENTFA  | Q93LK3 enterococcu |
| 29         | 436    | 20.7        | 211    | 2 Q8DVU5_STRMU  | Q8DVU5 streptococc |
| 30         | 420.5  | 20.0        | 544    | 2 Q840X3_STRMU  | Q840X3 streptococc |
| 31         | 418.5  | 19.9        | 544    | 2 Q840V8_STRMU  | Q840V8 streptococc |

|  |   |   |     |        |      |       |        |         |                        |
|--|---|---|-----|--------|------|-------|--------|---------|------------------------|
| RESULT 1   |   |   |     |        |      |       |        |         |                        |
| Q938V0_STRMU   | PRELIMINARY;  |   |     |        | PRT; | 431   | AA.    |         |                        |
| ID   | Q938V0  |   |     |        |      |       |        |         |                        |
| AC   | Q938V0;   |   |     |        |      |       |        |         |                        |
| DT   | 01-DEC-2001   | (TremBLrel. 19, Created)                |     |        |      |       |        |         |                        |
| DT   | 01-DEC-2001   | (TremBLrel. 19, Last sequence update)   |     |        |      |       |        |         |                        |
| DT   | 01-MAR-2004   | (TremBLrel. 26, Last annotation update) |     |        |      |       |        |         |                        |
| DE   | Glucan-binding protein B.   |   |     |        |      |       |        |         |                        |
| DE   | Streptococcus mutans.   |   |     |        |      |       |        |         |                        |
| OC   | Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;              |   |     |        |      |       |        |         |                        |
| OC   | Streptococcus   |   |     |        |      |       |        |         |                        |
| OX   | NCBI_TaxID=1309;  |   |     |        |      |       |        |         |                        |
| RN   | [1]   |   |     |        |      |       |        |         |                        |
| RP   | NUCLEOTIDE SEQUENCE.  |   |     |        |      |       |        |         |                        |
| RC   | STRAIN=5SM3;  |   |     |        |      |       |        |         |                        |
| RA   | Jin S., Duncan M.J., Taubman M.A., Smith D.J.;                        |   |     |        |      |       |        |         |                        |
| RT   | "Cloning of the gbpB gene from Streptococcus mutans.";                |   |     |        |      |       |        |         |                        |
| RL   | J. Dent. Res. 79:224-224(2000).                                       |   |     |        |      |       |        |         |                        |
| RN   | [2]   |   |     |        |      |       |        |         |                        |
| RP   | NUCLEOTIDE SEQUENCE.  |   |     |        |      |       |        |         |                        |
| RC   | STRAIN=5SM3;  |   |     |        |      |       |        |         |                        |
| RX   | MEDLINE=21481971; PubMed=11598068;                                    |   |     |        |      |       |        |         |                        |
| RX   | DOI=10.1128/IAI.69.11.6931-6941.2001;                                 |   |     |        |      |       |        |         |                        |
| RA   | Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,           |   |     |        |      |       |        |         |                        |
| RA   | Duncan M.J.;  |   |     |        |      |       |        |         |                        |
| RT   | "Cloning of the Streptococcus mutans gene encoding glucan binding     |   |     |        |      |       |        |         |                        |
| RT   | protein B and analysis of genetic diversity and protein production in |   |     |        |      |       |        |         |                        |
| RL   | clinical isolates.";  |   |     |        |      |       |        |         |                        |
| RL   | Infect. Immun. 69:6931-6941(2001).                                    |   |     |        |      |       |        |         |                        |
| DR   | EMBL; AY046414; AAK94504.1; -; Genomic_DNA.                           |   |     |        |      |       |        |         |                        |
| DR   | InterPro; IPR007921; CHAP.  |   |     |        |      |       |        |         |                        |
| DR   | InterPro; IPR009148; SIBA.  |   |     |        |      |       |        |         |                        |
| DR   | Pfam; PF05257; CHAP; 1.   |   |     |        |      |       |        |         |                        |
| DR   | PRINTS; PR01852; SIBAPROTEIN.   |   |     |        |      |       |        |         |                        |
| DR   | PROSITE; PS50911; CHAP; 1.  |   |     |        |      |       |        |         |                        |
| SQ   | SEQUENCE 431 AA; 44650 MW; 05D38D8B8C4609F CRC64;                     |   |     |        |      |       |        |         |                        |
| Query Match 100.0%; Score 2105; DB 2; Length 431;            |   |   |     |        |      |       |        |         |                        |
| Best Local Similarity 100.0%; Pred. No. 2.9e-86;             |   |   |     |        |      |       |        |         |                        |
| Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |   |   |     |        |      |       |        |         |                        |
| Qy   | 1   | MKKRILSAVLVSGVTLSAT                     | TL  | SAVKAD | FD   | DAQIA | SQDSK  | INNLTA  | QQAQAQVNTIQG 60        |
| Db   | 1   | MKKRILSAVLVSGVTLSAT                     | TL  | SAVKAD | FD   | DAQIA | SQDSK  | INNLTA  | QQAQAQVNTIQG 60        |
| Qy   | 61  | QVSALOTQAELOAENRLE                      | QA  | QSAT   | LG   | QOIQ  | ITLSSK | IVARNES | LKQOAFSAQNSAATSY 120   |
| Db   | 61  | QVSALOTQAELOAENRLE                      | QA  | QSAT   | LG   | QOIQ  | ITLSSK | IVARNES | LKQOAFSAQNSAATSY 120   |
| Qy   | 121   | INAIINSKSVSDAINRV                       | SAI | REVVS  | AN   | KMLQ  | QQQD   | KA      | AVFQKQQAINTVAANQE 180  |
| Db   | 121   | INAIINSKSVSDAINRV                       | SAI | REVVS  | AN   | KMLQ  | QQQD   | KA      | AVFQKQQAINTVAANQE 180  |
| Qy   | 181   | TTIAQNTNALNTQQA                         | LEA | AAQL   | NLQ  | LAEL  | TTA    | QDKAT   | LVQAQAAAEARQAQAAAE 240 |

Db 181 TIAQNTNALNTQQAQLERAAQLNLQALHTTAQDQKATLVAKKAAAEARQAQAAAE 240  
 QY 241 KAAAEAKALQEQAAQAAANNTQATDASDQAAADNTQAAQTGDSQSAQAQVNN 300  
 Db 241 KAAAEAKALQEQAAQAAANNTQATDASDQAAADNTQAAQTGDSQSAQAQVNN 300  
 QY 301 DOESTTATEAOPSASSASTAVTANTSSANTYPAGQCTGWKSLAPWYGNWNGGQWAA 360  
 Db 301 DOESTTATEAOPSASSASTAVTANTSSANTYPAGQCTGWKSLAPWYGNWNGGQWAA 360  
 QY 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTGQGGQIQVQEAANYAGNOSIGNYRGWF 420  
 Db 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTGQGGQIQVQEAANYAGNOSIGNYRGWF 420  
 QY 421 NPGSVSYIYPN 431  
 Db 421 NPGSVSYIYPN 431

RESULT 2  
 Q9AG98\_STRMU PRELIMINARY; PRT; 431 AA.  
 AC Q9AG98;  
 DT 01-JUN-2001 (TREMELrel. 17, Created)  
 DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)  
 DT 01-FEB-2005 (TREMELrel. 29, Last annotation update)  
 DE Immunodominant glycoprotein IDG-60 (Glucan-binding protein B).  
 GN Name=sagA;  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=GS-5;  
 RX MEDLINE=21153617; PubMed=11254612;  
 RX DOI=10.1128/IAI.69.4.2493-2501.2001;  
 RA Chia J.S., Lee Y.Y., Huang P.T., Chen J.Y.;  
 RA "Identification of stress-responsive genes in Streptococcus mutans by differential display reverse transcription-PCR."  
 RL Infect. Immun. 69:2493-2501(2001).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=GS-5;  
 RX MEDLINE=21481977; PubMed=11598074;  
 RX DOI=10.1128/IAI.69.11.6987-6998.2001;  
 RA Chia J.S., Chang L.Y., Shun C.T., Chang Y.Y., Chen J.Y.;  
 RA "A 60-kilodalton immunodominant glycoprotein is essential for cell wall integrity and the maintenance of cell shape in Streptococcus mutans."  
 RL Infect. Immun. 69:6987-6998(2001).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=GS-5;  
 RX MEDLINE=21481971; PubMed=11598068;  
 RX DOI=10.1128/IAI.69.11.6931-6941.2001;  
 RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,  
 RA Duncan M.J.;  
 RA "Cloning of the Streptococcus mutans gene encoding glucan binding protein B and analysis of genetic diversity and protein production in clinical isolates."  
 RT

RL Infect. Immun. 69:6931-6941(2001).  
 DR EMBL; AF338445; AAK08104.1; -; Genomic DNA.  
 DR EMBL; AY046411; AAK4501.1; -; Genomic DNA.  
 DR InterPro; IPR007921; CHAP.  
 DR InterPro; IPR009148; Siba.  
 DR Pfam; PF05257; CHAP; 1.  
 DR PRINTS; PR01852; SIBAPROTEIN.  
 DR PROSITE; PS50911; CHAP; 1.  
 SQ SEQUENCE 431 AA; 44592 MW; 3EBE21FC5E47232E CRC64;  
 Query Match 99.6%; Score 2096; DB 2; Length 431;  
 Best Local Similarity 99.5%; Pred. No. 7,3e-86;  
 Matches 429; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQSKINNLTAAQQAQAQVNTIQG 60  
 Db 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQSKINNLTAAQQAQAQVNTIQG 60  
 QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSAATSY 120  
 Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSAATSY 120  
 QY 121 INAIINSKVSDAINRVSAIREVWSANEKMLQQQPDKAAVEKQKQOENQAATNTVAANO 180  
 Db 121 INAIINSKVSDAINRVSAIREVWSANEKMLQQQPDKAAVEKQKQOENQAATNTVAANO 180  
 QY 181 TIAQNTNALNTQQAQLERAAQLNLQALHTTAQDQKATLVAKKAAAEARQAQAAAE 240  
 Db 181 TIAQNTNALNTQQAQLERAAQLNLQALHTTAQDQKATLVAKKAAAEARQAQAAAE 240  
 QY 241 KAAAEAKALQEQAAQAAANNTQATDASDQAAADNTQAAQTGDSQSAQAQVNN 300  
 Db 241 KAAAEAKALQEQAAQAAANNTQATDASDQAAADNTQAAQTGDSQSAQAQVNN 300  
 QY 301 DOESTTATEAOPSASSASTAVTANTSSANTYPAGQCTGWKSLAPWYGNWNGGQWAA 360  
 Db 301 DOESTTATEAOPSASSASTAVTANTSSANTYPAGQCTGWKSLAPWYGNWNGGQWAA 360  
 QY 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTGQGGQIQVQEAANYAGNOSIGNYRGWF 420  
 Db 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTGQGGQIQVQEAANYAGNOSIGNYRGWF 420  
 QY 421 NPGSVSYIYPN 431  
 Db 421 NPGSVSYIYPN 431  
 RESULT 3  
 Q938V3\_STRMU PRELIMINARY; PRT; 431 AA.  
 AC Q938V3;  
 DT 01-DEC-2001 (TREMELrel. 19, Created)  
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
 DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)  
 DE Glucan-binding protein B.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=SJ32;  
 RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;  
 RT "Cloning of the gbpB gene from Streptococcus mutans."  
 RL J. Dent. Res. 79:224-224(2000).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=SJ32;  
 RX MEDLINE=21481971; PubMed=11598068;  
 RX DOI=10.1128/IAI.69.11.6931-6941.2001;  
 RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,  
 RA Duncan M.J.;  
 RA "Cloning of the Streptococcus mutans gene encoding glucan binding

RT protein B and analysis of genetic diversity and protein production in

|  |   |
|--|---|
| RT   | Clinical isolates <sup>a</sup> :                                      |
| RL   | Infect. Immun. 89:6931-6941(2001).                                    |
| DR   | EMBL; AY046410; AAK34500.1; -; Genomic_DNA.                           |
| DR   | InterPro; IPR007921; CHAP.  |
| DR   | InterPro; IPR009148; Siba.  |
| DR   | Pfam; PF05257; CHAP; 1.   |
| DR   | PRINTS; PR01852; SIBAPROTEIN.   |
| DR   | PROSITE; PS50911; CHAP; 1.  |
| SQ   | SEQUENCE 431 AA; 44620 MW; 464FE3B563FB7E51 CRC64;                    |
| <br>   |   |
| Query Match  |   |
| Best Local Similarity 99.3%; Score 2092; DB 2; Length 431;   |   |
| Matches 428; Conservative 0; Mismatches 3; Indels 0; Gaps 0; |   |
| <br>   |   |
| Qy   | 1 MKKILSAVLVGGVTLSATTLSAVKADDFDAQIASODSKINNLTACOOAAQAQVNTITQG 60      |
| Dd   | 1 MKKILSAVLVGGVTLSATTLSAVKADDFDAQIASQDSKINNLTACOOAAQAQVNTITQG 60      |
| Qy   | 61 QVSALQTQQAEQLAENQRLEAQSATLGGQIQTLLSSKI VARNESLKQARSACKNSAATSY 120  |
| Dd   | 61 QVSALQTQQAEQLAENQRLEAQSATLGGQIQTLLSSKI VARNESLKQARSACKNSAATSY 120  |
| Qy   | 121 INAILNKSVSDAINRVSAIREVVSANEKMLQQOEQDKAAVEQEQENQAINTVAANQE 180     |
| Dd   | 121 INAILNKSVSDAINRVSAIREVVSANEKMLQQOEQDKAAVEQEQENQAINTVAANQE 180     |
| Qy   | 181 TIAQNTWALNTQQAELEAAQLNLQAE LTTADQKATLVAKAAEEAAROAQAAAAQAAEA 240   |
| Dd   | 181 TIAQNTWALNTQQAELEAAQLNLQAE LTTADQKATLVAKAAEEAAROAQAAAAQAAEA 240   |
| Qy   | 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAADNTQAAQTGDS TEQSAAQAVNNS 300 |
| Dd   | 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAADNTQAAQTGDS TEQSAAQAVNNS 300 |
| Qy   | 301 DQESTTATEAQPSASSASTAVVTANTSSANTYPAGQCTGWYKSLAPWGVNTWMNGGQWAA 360  |
| Dd   | 301 DQESTTATEAQPSASSASTAAVAANTSSANTYPAGQCTGWYKSLAPWGVNTWMNGGQWAA 360  |
| Qy   | 361 SAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQEANVAGNQSIGNYRGWF 420   |
| Dd   | 361 SAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQGGQIQVQEANVAGNQSIGNYRGWF 420   |
| Qy   | 421 NPGSVSVIYPN 431   |
| Dd   | 421 NPGSVSVIYPN 431   |

```

RESULT 4
QB8DMW3 STRMU
ID QB8DMW3 STRMU PRELIMINARY; PRT; 431 AA.
AC QB8DMW3;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Putative secreted antigen GbpB/Saga; putative peptidoglycan
DE hydrolase.
DN Name=gbpB; OrderedLocusNames=SMU.22;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
[1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=UAI59 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kanton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RA "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
RT pathogen."
RT Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
RL

```

|    |   |
|----|---|
| DR | EMBL; AE014855; AAN57811.1; -; Genomic_DNA.                             |
| DR | InterPro; IPR007921; CHAP.  |
| DR | InterPro; IPR009148; Siba.  |
| DR | Pfam; PF05257; CHAP; 1.   |
| DR | PRINTS; PR01852; SIBAPROTEIN.   |
| DR | PROSITE; PS0911; CHAP; 1.   |
| KW | Complete proteome.  |
| SQ | SEQUENCE 431 AA; 44620 MW; 2D1CA685248CCD3E CRC64;                      |
|    | Query Match 99.4%; Score 2092; DB 2; Length 431;                        |
|    | Best Local Similarity 99.3%; Fred.No. 1.le-85;                          |
|    | Matches 428; Conservative 0; Mismatches 3; Indels 0; Gaps 0             |
| Qy | 1 MKKRLSAVLVSGVTLSATTLSAVKADDDFAQIASQDSKINNLTAAQQAAAQAQVNTIOG 60        |
| Dd | 1 MKKRLSAVLVSGVTLSATTLSAVKADDDFAQIASQDSKINNLTAAQQAAAQAQVNTIOG 60        |
| Qy | 61 QVSALQTQQAELQAELOENRLEAQSATLGOOIOTLLSSKIVARNESLKQARSQAOKSNAATSY 120  |
| Dd | 61 QVSALQTQQAELQAELOENRLEAQSATLGOOIOTLLSSKIVARNESLKQARSQAOKSNAATSY 120  |
| Qy | 121 INAIINSKSVSDAINRVSAIREVVSANEKMLQQEQDKAAAEVKQEQENQAALINTVAANCE 180   |
| Dd | 121 INAIINSKSVSDAINRVSAIREVVSANEKMLQQEQDKAAAEVKQEQENQAALINTVAANCE 180   |
| Qy | 181 TIAQNTNALNTQQAQLAAQLNLQBELTTAQDQKATLVAKKAABEAARQAQAAAAQA 240        |
| Dd | 181 TIAQNTNALNTQQAQLAAQLNLQBELTTAQDQKATLVAKKAABEAARQAQAAAAQA 240        |
| Qy | 241 KAAAEAKALQEQAAAQAQAAANNNTQATDASDQQAADAADNTQAAQTGDSTEQSAAQAQVNNS 300 |
| Dd | 241 KAAAEAKALQEQAAAQAQAAANNNTQATDVSDQQAADAADNTQAAQTGDSTEQSAAQAQVNNS 300 |
| Qy | 301 DQESTTATEAQPSASSASTAVVTANTSSANTYPACQCTGWUKSLAPWVGNYNGNGQOWAA 360    |
| Dd | 301 DQESTTATEAQPSASSASTAAVAANTSSANTYPACQCTGWVKSLAPWVGNYNGNGQOWAA 360    |
| Qy | 361 SAAAGYRVGSTPISAGAVAVNDGGYGHVAVYTGVQGGOIQVOEANYAGNQSIGINYRGWF 420    |
| Dd | 361 SAAAGYRVGSTPISAGAVAVNDGGYGHVAVYTGVQGGOIQVOEANYAGNQSIGINYRGWF 420    |
| Qy | 421 NPGSVSYIYNP 431   |
| Dd | 421 NPGSVSYIYNP 431   |

```

RESULT 5
Q93BV2_STRMU
ID Q93BV2_STRMU PRELIMINARY; PRT; 432 AA.
AC Q93BV2;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=1309;
OX [1]
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=15JP2;
RX MEDLINE=21481971;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
EL J. Dent. Res. 79:224-224(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=15JP2;
RX MEDLINE=21481971; PubMed=11598068;
RA DOI=10.1128/TAI.69.11.6931-6941.2001;
RT Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in

```

```
RT clinical isolates.";
RL Infect Immun. 69:6931-6941(2001).
DR EMBL; AY046412; AAK94502.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44648 MW; E769B2504AEE50E9 CRC64;

Query Match      98.6%; Score 2075.5; DB 2; Length 432;
Best Local Similarity 98.6%; Pred. No. 5.9e-85;
Matches 426; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQOAAAQAVNTTIG 60
Db 1 MKKRILSAVLVSGVTLSSATTLISAKADDFDAQIASQDSKINNLTAAQOAAAQAVNTTIG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNAAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNAAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAAINTVAANQE 180
QY 181 TTAQNTNALNTQOQAELOAENORLEAQLNQLAELTTAQQOKATLVAQKAAAEARQAAAAQAAEA 240
Db 181 TTAQNTNALNTQOQAELOAENORLEAQLNQLAELTTAQQOKATLVAQKAAAEARQAAAAQAAEA 240
QY 241 KAAAEKALQEQAAQAAAA--NNNTQATDASDQAAAAADNTQAAOTGDSSTEQSAAQAVNN 299
Db 241 KAAAEKALQEQAAQAAAA--NNNTQATDASDQAAAAADNTQAAOTGDSSTEQSAAQAVNN 300
QY 300 SDQESTTATEAQPSSASSASTAVVTANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 359
Db 301 SDQESTTATAAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 360
QY 360 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANTAGNOSIGNYRGW 419
Db 361 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANTAGNOSIGNYRGW 420
QY 420 FNPGSVSYIYPN 431
Db 421 FNPGSVSYIYPN 432

RESULT 6
Q938V1_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gopB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/JAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding

RT protein B and analysis of genetic diversity and protein production in
RL Clinical isolates.";
RL Infect Immun. 69:6931-6941(2001).
DR EMBL; AY046413; AAK94503.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44652 MW; 3F88ECB9A1F3BE4F CRC64;

Query Match      98.2%; Score 2066.5; DB 2; Length 432;
Best Local Similarity 98.4%; Pred. No. 1.5e-84;
Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQOAAAQAVNTTIG 60
Db 1 MKKRILSAVLVSGVTLSSATTLISAVKADDFDAQIASQDSKINNLTAAQOAAAQAVNTTIG 60
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNAAATSY 120
Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSQKSNAAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDKAAVEQKQENQAAINTVAANQE 180
QY 181 TTAQNTNALNTQOQAELOAENORLEAQLNQLAELTTAQQOKATLVAQKAAAEARQAAAAQAAEA 240
Db 181 TTAQNTNALNTQOQAELOAENORLEAQLNQLAELTTAQQOKATLVAQKAAAEARQAAAAQAAEA 240
QY 241 KAAAEKALQEQAAQAAAA--NNNTQATDASDQAAAAADNTQAAOTGDSSTEQSAAQAVNN 299
Db 241 KAAAEKALQEQAAQAAAA--NNNTQATDASDQAAAAADNTQAAOTGDSSTEQSAAQAVNN 300
QY 300 SDQESTTATEAQPSSASSASTAVVTANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 359
Db 301 SDQESTTATAAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWGGGOWA 360
QY 360 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANTAGNOSIGNYRGW 419
Db 361 ASAAAAGRVGTPSAGAVAVVNDGGYGHVAVVTGVGGQIQVQAEYANTAGNOSIGNYRGW 420
QY 420 FNPGSVSYIYPN 431
Db 421 FNPGSVSYIYPN 432

RESULT 7
Q9AKA4_STRAG PRELIMINARY; PRT; 447 AA.
AC Q9AKA4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PcsB protein precursor.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=6313;
RX MEDLINE=21101799; PubMed=11157929;
RX DOI=10.1128/JB.183.4.1175-1183.2001;
RA Reinscheid D.J., Gottschalk B., Schubert A., Eikmanns B.J.,
RA Chhatwal G.S.;
RT "Identification and molecular analysis of PcsB, a protein required for
RT cell wall separation of group B streptococcus.";
RL J. Bacteriol. 183:1175-1183(2001).
DR EMBL; AJ277292; CAC28144.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
```



|          |  |
|----------|--|
| DR       | InterPro; IPR009148; Siba.   |
| DR       | Pfam; PF05257; CHAP; 1.  |
| DR       | PRINTS; PRO1852; SIBAPROTEIN.  |
| DR       | PROSITE; PS0911; CHAP; 1.  |
| KW       | Signal.  |
| FT       | SIGNAL 1 25 Potential.   |
| FT       | CHAIN 26 447 Pcsb protein.   |
| FT       | SEQUENCE 447 AA; 46681 MW; F4DBI4BOA5F962C8 CRC64;                       |
| QY       | Query Match 58.8%; Score 1238; DB 2; Length 447;                         |
| QY       | Best Local Similarity 60.7%; Pred. No. 1.1e-47;                          |
| QY       | Matches 280; Conservative 43; Mismatches 94; Indels 44; Gaps 10;         |
| QY       | 1 MKKRILSAVLVSGVTLSSATTLSSAVKADDDFAQIASODSKINNLTAAQQQAAQAQVNTIQG 60      |
| QY       | 1 MKKRILSAVLVSGVTLTGTAAV--TVNADDDFSKIAATDSVINTLSGQAAQAAQNVTAIKG 58       |
| QY       | 61 QVSAQTQOAELOEAOENORLEAQAATLGGQOTQOTLSSKIVARNESLKKQAARSQAQSNAAATSY 120 |
| QY       | 59 QVGALESQQSELEAQAQNAQLSAVSOQLGOEIQTLNSKNIVARNESLUKKQVRSQAQGN-LTNY 117  |
| QY       | 121 INAIINSKSVSDAINRYSALREVVSANEKMLQQEQDQKAAVEQKQENQQAINTVAAAOE 180      |
| QY       | 118 INTILNSKSVSDAVNRVVAIREVVSANEKMLAQEADKAALEAKQIENQAINTVAAQKQ 177       |
| QY       | 181 TIAQNTNALNTQQAQLEAAQNLQAEILTTAQDQKATLVAQKAAAEAEAAQAAAAQAAEA 240      |
| QY       | 178 AIENKKAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQAEAAARKAAEAQAAAAEA 237    |
| QY       | 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQQAQAAADNTQAAQTGDSSTE--QSAQAQVN 298  |
| QY       | 238 KAQAQEAQAQAESVAKAQAQAA-----QVESATAPTETVQTQPRTEIKPSNLTAITS 286        |
| QY       | 299 NSDQESTTATEA-----OPS-----ASSASTAVVTA-----NTSSANTYPAGQ 336            |
| QY       | 287 SATTVATTTATATATNEPKVTQPSVVTKAVEAPKAVSVSTPRAVSKPVRSYDSSNTYPMGQ 346    |
| QY       | 337 CTWGVKSLAPWGVNMGWGQGWASAAAAAGYRGVSTPSAGAVAYW--NDGGYGHVAYVTG 394      |
| QY       | 347 CTWGAQSMASWGVNMGWGNQWASARAAGYSGVTTPRVGAVAVWPYDGGYGHVAVVTS 406        |
| QY       | 395 V-QGGQIQVQAEANYAGNQSGTNYRGWNP---GSVSYIYPN 431                        |
| QY       | 407 VANNSSIQVMESNYAGNMSIGNYRGSEFNPSASGSVYIYPN 447                        |
| RESULT 8 |  |
| ID       | Q8E2H1_STRAS PRELIMINARY; PRT; 447 AA.                                   |
| AC       | Q8E2H1;  |
| DT       | 01-MAR-2003 (TrEMBLrel. 23, Created)                                     |
| DT       | 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)                        |
| DT       | 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)                      |
| DE       | Pcsb protein.  |
| GN       | Names=pscb; OrderedLocusNames=SAG0017;                                   |
| OS       | Streptococcus agalactiae (serotype V).                                   |
| OC       | Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;                 |
| OC       | Streptococcus.   |
| OX       | NCBI_TaxID=216466;   |
| RP       | [1]  |
| RP       | NUCLEOTIDE SEQUENCE.   |
| RC       | STRAIN=2603 V/R / Serotype V;  |
| RC       | MEDLINE=2222298; PubMed=12200547; DOI=10.1073/pnas.182380799;            |
| RA       | Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.N.,    |
| RA       | Wessels L.R., Paulsen I.T., Nelson K.B., Margarit I., Read T.D.,         |
| RA       | Madoff L.C., Wolf A.M., Beaman M.J., Brinkac L.M., Daugherty S.C.,       |
| RA       | DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,            |
| RA       | Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,          |
| RA       | Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarcellii M., Mora M.,  |
| RA       | Iacoboni E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,   |
| RA       | Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,           |
| RA       | Fraser C.M.;   |
| RT       | "Complete genome sequence and comparative genomic analysis of an         |

emerging human pathogen, serotype V *Streptococcus agalactiae*.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).  
 RL EMBL; AE014192; AM98925.1; -; Genomic\_DNA.  
 DR TIGR; SAG0017; -;  
 DR InterPro; IPR007921; CHAP.  
 DR InterPro; IPR009148; SIBA.  
 DR Pfam; PF05257; CHAP; 1.  
 DR PRINTS; PR01852; SIBAPROTEIN.  
 DR PROSITE; PS00911; CHAP; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 58.8%; Score 1238; DB 2; Length 447;  
 Best Local Similarity 60.7%; Pred. No. 1.1e-47;  
 Matches 280; Conservative 43; Mismatches 94; Indels 44; Gaps 10

QY 1 MKKRIISAVLVSGVTLSATTLSAVKADDPDAQIASODSKINNLTIAQOCAAQAVNTIQG 60  
 DB 1 MKKRIISAVLVSGVTLTAAV--TVNADDDFSKLAATDSVINTLSGQAAANQVTAIKG 58  
 QY 61 QVSALQTQQAELQAEORLEAQSATLGGQIQITLSKSVARNESLKQARSQAQSNAAATSY 120  
 DB 59 QVGALESQQSELEAQNQALEAVSQQLGQEIQTLSNKIVARNESLKQVRSQAQGN-LTNY 117  
 QY 121 INAIINSKSVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQEQENQAINTVAAHQE 180  
 DB 118 INTILNSKSVSDAVNRVVAIREVVSANEKMLAQOQADKAALQAEKQIENQNAINVAAANKQ 177  
 QY 181 TIAQNTNALNTQQAELAAQNLQAEITTAQDOKATLVAQKAAEEAARQAAAAQAAAEA 240  
 DB 178 AIENKKAALATQRAQLAQAQLSAQLTTVQNEKASLIQAKAAEEAARKAAEQAAAEA 237  
 QY 241 KAAAEAKALQEAQAAQAAAAANNNTQATDASDQAAAAADNTQAAQTGDSTE--QSAQAQVN 298  
 DB 238 KAQAQEAQAQESVAKAQAAA-----QVESATAPTETVQTPRTEIKPSNLTATS 286  
 QY 299 NSDQESTATEA-----QPS-----ASSASTAVVTA-----NTSSANTYPAGQ 336  
 DB 287 SATTVAITTTATATNEPKVTQPSVVTKAVEAPKAVVSTTPRAVSPVRSYDSSNTYPMGQ 346  
 QY 337 CTWGVKSLAPWGNVWNGGOWAASAAAAGYRGVSTPSAGAVAVW--NDGGVGHVAVYTG 394  
 DB 347 CTWGAKSMAWGNVWGNQANQWASARAAGYSGVTTPRVGAVAVWVPYDGGVGHVAVVTS 406  
 QY 395 V-QGGQIQIQEANYAGNQSIGNRYGWFNP---GSVSYIYPN 431  
 DB 407 VANNSSIQWESNYAGNMSIGNRYGSENPSPASGSVYIYPN 447

RESULT 9  
 Q8E7X9\_STRA3  
 ID Q8E7X9\_STRA3 PRELIMINARY; PRT; 447 AA.  
 AC Q8E7X9;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE PCsb protein.  
 GN Name=pcsb; OrderedLocusNames=gb0016;  
 OS Streptococcus agalactiae (serotype III).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 RN NCBI\_TaxID=216495;  
 RX [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=NEW316 / Serotype III;  
 RX MEDLINE=22422508; PubMed=12354221;  
 RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Prangeul L.,  
 RA Meadek T., Zouine M., Couve E., Lalloui L., Poyart C., Trieu-Cuot P.,  
 RA Kunst F.;  
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
 invasive neonatal disease.";  
 RL Mol. Microbiol. 45:1499-1513 (2002).  
 DR EMBL; AL766843; CAD45661.1; -; Genomic DNA.



```

RT M6 strain."
RL J. Infect. Dis. 190:727-738 (2004).
DR EMBL; CP000003; AAT86152.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match      51.9%; Score 1092.5; DB 2; Length 398;
Best Local Similarity 53.7%; Pred. No. 3e-41;
Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIIG 60
DB 1 MKKRILSAVLVSGVTLGAATT---VCAEDLSKIAKQDSIIINLTTEQKAQNVSAALQA 57
QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQITLSSKIVARNESLKQQAARSQAQNSAATSY 120
DB 58 QVSSLQSEODKLTARNTLEALSKEFEQIKALTSQIVARNEKLNQQAARSAYKNNETSGY 117
QY 121 INAININSKVSDAINRVSAIREVSVANEKMLQQEQDKAAVEQKQOENQAQNTVAANOGE 180
DB 118 INALLNSKISIDVNVRLVAINRAVSAKLEEQKADKVSLEEKQAANTQTAINTIAANMA 177
QY 181 TTAQNTNALNTQQAOLEAAQLNLOAELTTAODOKATLVAQKAAAEAEARQAQAAAAEA 240
DB 178 MAEENQNTLRTQANLEAATANLALQASATEDKANLVAQKEAAEKAAAEALAEQAQAKV 237
QY 241 KAAAEAKALQEQAAQAQAANNTQATDASDQQAADNTQAQTDGDSQSAQAQVNNNS 300
DB 238 KAEQEA---AQQAASVEAAKSAITPAQA-----TPAAQSSNAIEPAALTA--- 280
QY 301 DQESTTATEAQSASASTAVVTANTSSANTYPAGQCTGWGKSLAPWVGNVWNGGOWAA 360
DB 281 ----PAAPSARQTS-----YDSSNTYPVQCCTGWGAKSLAPWAGNNGGOWAY 326
QY 361 SAAAAGYRGVSTPSAGAVAVWNDGGYGHVAVYTVGQ--GGQIQVOEANYAGNOSIGNYRGW 419
DB 327 SAQAAGYRGVSTPMVGAIAVWNDGGYGHVAVVVEQSSASSIRVMESNYSGROYIADHRGW 386
QY 420 FNPQSVSYIYPN 431
DB 387 FNPTGVTFIYPH 398

RESULT 13
QBP318_STRP3
ID QBP318_STRP3 PRELIMINARY; PRT; 398 AA.
AC QBP318; Q7CFL7;
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 13-SEP-2005 (TremBLrel. 31, Last annotation update)
DE Putative secreted protein.
GN OrderedLocusNames=SP0015; Spym3_0014;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301448;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SSI-1 / Serotype M3;
RX MEDLINE=22683278; PubMed=12799345; DOI=10.1101/gr.1096703;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution."
RL Genome Res. 13:1042-1055(2003).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MGAS315 / Serotype M3;

M6 strain."
J. Infect. Dis. 190:727-738 (2004).
EMBL; CP000003; AAT86152.1; -; Genomic_DNA.
InterPro; IPR007921; CHAP.
Pfam; PF05257; CHAP; 1.
PRINTS; PR01852; SIBAPROTEIN.
PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match      51.9%; Score 1092.5; DB 2; Length 398;
Best Local Similarity 53.7%; Pred. No. 3e-41;
Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIIG 60
DB 1 MKKRILSAVLVSGVTLGAATT---VCAEDLSKIAKQDSIIINLTTEQKAQNVSAALQA 57
QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIQITLSSKIVARNESLKQQAARSQAQNSAATSY 120
DB 58 QVSSLQSEODKLTARNTLEALSKEFEQIKALTSQIVARNEKLNQQAARSAYKNNETSGY 117
QY 121 INAININSKVSDAINRVSAIREVSVANEKMLQQEQDKAAVEQKQOENQAQNTVAANOGE 180
DB 118 INALLNSKISIDVNVRLVAINRAVSAKLEEQKADKVSLEEKQAANTQTAINTIAANMA 177
QY 181 TTAQNTNALNTQQAOLEAAQLNLOAELTTAODOKATLVAQKAAAEAEARQAQAAAAEA 240
DB 178 MAEENQNTLRTQANLEAATANLALQASATEDKANLVAQKEAAEKAAAEALAEQAQAKV 237
QY 241 KAAAEAKALQEQAAQAQAANNTQATDASDQQAADNTQAQTDGDSQSAQAQVNNNS 300
DB 238 KAEQEA---AQQAASVEAAKSAITPAQA-----TPAAQSSNAIEPAALTA--- 280
QY 301 DQESTTATEAQSASASTAVVTANTSSANTYPAGQCTGWGKSLAPWVGNVWNGGOWAA 360
DB 281 ----PAAPSARQTS-----YDSSNTYPVQCCTGWGAKSLAPWAGNNGGOWAY 326
QY 361 SAAAAGYRGVSTPSAGAVAVWNDGGYGHVAVYTVGQ--GGQIQVOEANYAGNOSIGNYRGW 419
DB 327 SAQAAGYRGVSTPMVGAIAVWNDGGYGHVAVVVEQSSASSIRVMESNYSGROYIADHRGW 386
QY 420 FNPQSVSYIYPN 431
DB 387 FNPTGVTFIYPH 398

RESULT 12
Q7CNO7_STRP8
ID Q7CNO7_STRP8 PRELIMINARY; PRT; 398 AA.
AC Q7CNO7;
DT 05-JUL-2004 (TremBLrel. 27, Created)
DT 05-JUL-2004 (TremBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TremBLrel. 27, Last annotation update)
DE Putative secreted protein.
GN OrderedLocusNames=spym18_0020;
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301451;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108; DOI=10.1073/pnas.062526099;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).

```

```
RX MEDLINE=22133808; PubMed=12122206; DOI=10.1073/pnas.152298499;
RA Berses S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
DR EMBL; BA000034; BAC63110.1; -; Genomic DNA.
DR EMBL; AF014136; AM78621.1; -; Genomic DNA.
DR InterPro; IPR009148; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match 51.9%; Score 1092.5; DB 2; Length 398;
Best Local Similarity 53.7%; Pred. No. 36-41;
Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAKADDFDAQIASQDSKINNLTAAQOAAQAVNTIQG 60
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKKRILSAVLVSGVTLGAATT---VGAEDLSFKIAQDSIIISLTTTEKKAQNOVSALQA 57
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKNAATSY 120
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 58 QVSSIQSEQDKLTARNTLEALSKEPQEI KALTSQIVARNESLKQOARSAQKNETSGY 117
QY 121 INAINSKSVSDAIRVSAIREVSAANEKMLQOQODKAAVEKQOENQAINTVAANQE 180
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 118 INALLNSKISDVAVRLVAIRVAVSANAKLLEQKADKRVSLBEKQAANQTALNTAANNA 177
QY 181 TTAQNTNALNTQOALEAAQLNQLAELTTAODKATLVAQKAAAEAEARQAAAEAA 240
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 178 MAEENQNTLRTQANLEAATANLALQASATEDKANLVAQKAAAEAEAAEAQQAQKV 237
QY 241 KAAABAKALOEAQAAQAAANNNTQATDASDQOAAADNTQAAQTGDSQTEQAAQAVNNS 300
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 238 KAQEQA---AQQAASVERAKSAITPAQA-----TPAQSNAIEPAALTA---- 280
QY 301 DOESTTATEAOPSASSASTAVTANTSSANTYPAGCTGWKSLAPWVGNVWNGGOWAA 360
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 281 ----PAAPSARPOTS-----YDSNTYTPVGCTGWAKSLAPWAGNNGGOWAY 326
QY 361 SAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTGVO-GGQIQVQEAANYAGNQSIGNYRGW 419
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 327 SAQAAGYRTGSTPMVGAIAVNDGGYGHVAVVVEVQSASSIRVMESNYSGRQVIADHRCW 386
QY 420 FNPGSVSYIYPN 431
DQ |||||:|||||:
Db 387 FNPGTGVTIYPH 398

RESULT 14
QSM212_STR11
ID QSM212_STR11 PRELIMINARY; PRT; 485 AA.
AC QSM212;
DT 01-FEB-2005 (TremBLrel. 29, Created)
DT 01-FEB-2005 (TremBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TremBLrel. 29, Last annotation update)
DE Glucan binding protein.
GN Name=pcdB; OrderedLocusNames=str0022;
OS Streptococcus thermophilus (strain CNRZ 1066).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=299768;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15543133; DOI=10.1038/nbt1034;
RA Bolotin A., Quinquis B., Renault P., Sorokin A., Ehrlich S.D.,
```

```
RA Kulakauskas S., Lapidus A., Goltsman E., Mazur M., Pusch G.D.,
RA Fonstein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D.,
RA Ngui K., Masuy D., Hancy F., Burteau S., Boutry M., Delcour J.,
RA Goffeau A., Hols P.;
RT "Complete sequence and comparative genome analysis of the dairy
RT bacterium Streptococcus thermophilus.";
RL Nat. Biotechnol. 22:1554-1558(2004).
DR EMBL; CP000024; AAV61641.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 485 AA; 49165 MW; 8665E4B1FC6CF7CA CRC64;

Query Match 51.7%; Score 1087.5; DB 2; Length 485;
Best Local Similarity 50.8%; Pred. No. 66-41;
Matches 248; Conservative 67; Mismatches 112; Indels 61; Gaps 12;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAKADDFDAQIASQDSKINNLTAAQOAAQAVNTIQG 60
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKKRILSAVLVSGVTLSSA---ASVHAEDYDSQIAATNTNAISNLASQOEAAQAVATQIS 57
QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKNAATSY 120
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 58 QVSTLRTQKTELEAKVAELEKVSADLESEIQELSKIVARQDSLAKQARSAQONTATSY 117
QY 121 INAINSKSVSDAIRVSAIREVSAANEKMLQOQODKAAVEKQOENQAINTVAANQE 180
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 118 INSILNSKISSEAITRITAISKVVTANNLLTKQESDQKELAAKQENQAINTTAAANKS 177
QY 181 TTAQNTNALNTQOALEAAQLNQLAELTTAODKATLVAQKAAAEAEARQAAAEAA 236
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 178 ELETTEAGLTTCQAELEAAQVTLAAELATAQNEKTSLSVSAKSTAESVAASTAASVAQSOA 237
QY 237 AAEAKAAAEAKALQEA---AQOAAANNNTQA-----TDASDQOAA---AADNTQAAQTGD 287
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 238 TAESNTAQVVASSEATSVASSEVATSEAVQOPSETPVSEISTASEAQPASETSE 297
QY 288 STEQSAQAQVNN-----SDQESTTAT-----EAQPSASSASTAVVVTANT 326
DQ |||||:|||||:
Db 298 VQESAAPAVSEAPASVAPVATSEAAPATSEAPASVAPVATSEAPASVAPVAPV 357
QY 327 SSA-----NTYPAGCTGWKSLAPWVGNVWNGGOWAAASAAAAGYRVGS 371
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 358 SEAAPAAETHKVSAASTNTYTPVGCTGWKSLAPWAGNNGGOWAY 417
QY 372 TPSAGAVAVW-ND-GGYGHVAVYTGVOG-GQIQVQEAANYAGNQSIGNYRGWFNP----- 422
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 418 TVAGAIAPWPDGGYGHVAVYTVTSAGANSIQVMESYAGNWSISNYRGTFDPTSSAHG 477
QY 423 GSVSYIYP 430
DQ |||||:
Db 478 GSVFYIYP 485

RESULT 15
Q9A128_STRPY
ID Q9A128_STRPY PRELIMINARY; PRT; 398 AA.
AC Q9A128; O7BHS9;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 10-MAY-2005 (TremBLrel. 30, Last annotation update)
DE Secreted protein Siba precursor.
GN OrderedLocusNames=SPY0019;
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
```

```

RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Pagan P.K., Reinscheid D., Gottschalk B., Chhatwal G.S.;
RT "Identification and characterization of a novel secreted protein from
RT group A streptococcus.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF006474; AAK33158.1; -; Genomic DNA.
DR EMBL; AF319999; AAL73135.1; -; Genomic DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome; Signal.
FT SIGNAL 1 23 Potential.
SQ SEQUENCE 398 AA; 41899 MW; 28A9B3F7195E969B CRC64;

Query Match 51.6%; Score 1085.5; DB 2; Length 398;
Best Local Similarity 53.7%; Pred. No. 6.1e-41;
Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSSATLISAVKADFDIAIASODSKINNLTAAQQAQAQVNTIQG 60
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MKKRILSAVLVSGVTLGAATT--VGAEDLSTKIAKQDSIISNLTTEQKAAQNVSAQQA 57
QY 61 QVSALQTOQAEIQAENORLEAQSATLIGQOIOTLSSKIVARNESLKKQARSQAQSNATSY 120
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58 QVSSILQSEQDKLTARNTLEALSKEPEQIKALTSQIVARNEKLNQAARSAYKNNETSGY 117
QY 121 INAIINSKSVDAINRVSAIREVVSANEKMLQQQSDKAAVEQKQOENQAINTVAANOE 180
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
118 INALLNKSISDVNRLVAINRAVSAKLEEQKADKVSLEEKQAANQTAINTIAANMA 177
QY 181 TTAQNTNALNTOQAOLEAQLNLAELTTAQDOKATLVAQKAAAEAAEQAAAQAQAAEA 240
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
178 MAENONTLRTQANLVAAATNLAQLASATEDKANLVAQKAAAEKAAEAQAQAAKV 237
QY 241 KAAAEAKALQEQAAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSTEQSAQAQAVNNS 300
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 KA-----QEQAA-----QQAASVEAAKSAITPAPQATPAAQSSNAI 273
QY 301 DQESTTATEAQPASASASTAVVTANTSSANTYPAGCTWGVKSLAPWGVNNGGQWAA 360
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
274 EPAALTA----PAAPSAQP---QTSYDSSNTYPVGQCTWGAKS LAPWAGNNNGGQWY 326
QY 361 SAAAGYRVGSTPSAGAVAVNDGGYGVHYVTGQV-GGQIOVQEVANVAGNQSIGNYRCW 419
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
327 SAQAAGYRTGSPVMGAIAMVNDGGYGVHVVVEVQSSIRVMSNYSYSGROYIADHRGW 386
QY 420 FNPGSVSYIYPN 431
DB |||||:|||||:
387 FNPTGVTFIYPH 398
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:09:37 ; Search time 29.9537 Seconds  
(without alignments)  
1189.611 Million cell updates/sec

Title: US-10-797-821-33

Perfect score: 2105

Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNRYGMFNGSGVYIYPN 431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5\_COMB.pep.\*  
2: /cgn2\_6/prodata/1/1aa/6\_COMB.pep.\*  
3: /cgn2\_6/prodata/1/1aa/H\_COMB.pep.\*  
4: /cgn2\_6/prodata/1/1aa/ECTUS\_COMB.pep.\*  
5: /cgn2\_6/prodata/1/1aa/RS\_COMB.pep.\*  
6: /cgn2\_6/prodata/1/1aa/backfilesl.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID                     | Description       |
|------------|--------|-------------|--------|------------------------|-------------------|
| 1          | 1057.5 | 50.2        | 399    | 2 US-09-107-433-3230   | Sequence 3230, Ap |
| 2          | 1056.5 | 50.2        | 392    | 2 US-09-583-110-4374   | Sequence 4374, Ap |
| 3          | 671    | 31.9        | 461    | 1 US-08-186-222-2      | Sequence 2, Appl  |
| 4          | 594    | 28.2        | 525    | 2 US-09-107-532A-5095  | Sequence 5095, Ap |
| 5          | 562    | 26.7        | 210    | 2 US-09-222-938A-67    | Sequence 67, Appl |
| 6          | 526.5  | 25.0        | 497    | 2 US-09-134-000C-5990  | Sequence 5990, Ap |
| 7          | 507    | 24.1        | 449    | 2 US-09-071-035-482    | Sequence 482, App |
| 8          | 507    | 24.1        | 449    | 2 US-10-206-576-482    | Sequence 482, App |
| 9          | 507    | 24.1        | 450    | 2 US-09-134-000C-5714  | Sequence 5714, Ap |
| 10         | 458    | 21.8        | 422    | 2 US-09-071-035-484    | Sequence 484, App |
| 11         | 458    | 21.8        | 422    | 2 US-10-206-576-484    | Sequence 484, App |
| 12         | 278    | 13.2        | 469    | 2 US-09-489-039A-13565 | Sequence 13565, A |
| 13         | 235    | 11.2        | 257    | 2 US-09-710-279-3244   | Sequence 3244, Ap |
| 14         | 235    | 11.2        | 264    | 2 US-09-134-001C-5035  | Sequence 5035, Ap |
| 15         | 235    | 11.2        | 267    | 2 US-09-134-001C-4539  | Sequence 4539, Ap |
| 16         | 230.5  | 11.0        | 2310   | 2 US-09-874-923-120    | Sequence 120, App |
| 17         | 228.5  | 10.9        | 477    | 2 US-09-902-540-11649  | Sequence 11649, A |
| 18         | 225.5  | 10.7        | 1236   | 2 US-09-769-787-109    | Sequence 109, App |
| 19         | 220    | 10.5        | 468    | 2 US-09-328-352-6321   | Sequence 6321, Ap |
| 20         | 218.5  | 10.4        | 610    | 2 US-09-336-447A-11    | Sequence 11, Appl |
| 21         | 218.5  | 10.4        | 610    | 2 US-09-952-267B-11    | Sequence 11, Appl |
| 22         | 215    | 10.2        | 149    | 2 US-09-710-279-1682   | Sequence 1682, Ap |
| 23         | 215    | 10.2        | 157    | 2 US-09-710-279-2870   | Sequence 2870, Ap |
| 24         | 215    | 10.2        | 270    | 2 US-09-134-001C-5441  | Sequence 5441, Ap |
| 25         | 210    | 10.0        | 1566   | 1 US-08-687-956A-23    | Sequence 23, Appl |
| 26         | 209.5  | 10.0        | 440    | 2 US-08-302-756B-35    | Sequence 35, Appl |
| 27         | 209    | 9.9         | 266    | 2 US-09-134-001C-5453  | Sequence 5453, Ap |

ALIGNMENTS

RESULT 1

US-09-107-433-3230  
; Sequence 3230, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
; THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: <Unknown>  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: <Unknown>  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,433  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/ 085131  
; FILING DATE: May 12, 1998  
; APPLICATION NUMBER: 60/051553  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arinello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-011  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 3230:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 399 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...399  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3230:

28 207.5 9.9 639 2 US-09-902-540-14908 Sequence 14908, A  
29 206 9.8 679 2 US-09-489-039A-12307 Sequence 12307, A  
30 205.5 9.8 1070 2 US-09-902-540-13861 Sequence 13861, A  
31 199 9.5 1561 2 US-08-894-017-23 Sequence 23, Appl  
32 199 9.5 1561 2 US-09-456-474-23 Sequence 23, Appl  
33 199 9.5 1565 6 5352450-2 Patent No. 5352450  
34 198.5 9.4 2101 1 US-08-466-390-4 Sequence 4, Appl  
35 198.5 9.4 2101 1 US-08-470-950-4 Sequence 4, Appl  
36 198.5 9.4 2101 1 US-08-467-781-4 Sequence 4, Appl  
37 198.5 9.4 2101 1 US-08-195-487-4 Sequence 4, Appl  
38 198.5 9.4 2101 1 US-08-483-924-4 Sequence 4, Appl  
39 198.5 9.4 2101 2 US-09-452-294-1 Sequence 1, Appl  
40 198.5 9.4 2101 2 US-09-296-662-32 Sequence 32, Appl  
41 198.5 9.4 2101 4 PCT-US93-06160-4 Sequence 4, Appl  
42 198 9.4 655 2 US-09-902-540-10005 Sequence 10005, A  
43 198 9.4 1024 2 US-09-270-767-44973 Sequence 44973, A  
44 197.5 9.4 2107 2 US-09-949-016-7646 Sequence 7646, Ap  
45 197.5 9.4 2107 2 US-09-949-016-7647 Sequence 7647, Ap

US-09-107-433-3230

Query Match 50.2%; Score 1057.5; DB 2; Length 399;  
Best Local Similarity 50.6%; Pred. No. 9.4e-70;  
Matches 220; Conservative 73; Mismatches 91; Indels 51; Gaps 5;

QY 1 MKKRILSAVLGVLTSSATTLSAVKADDFDAQIASQDSKINNLTAQQAQAQVNTTIG 60  
DB 8 MKKRILASLLSTVWVSQVAVLTTAAHETTTDDKIAAQDNKISNLTAQQAQAQVNTTIG 67

QY 61 QVSALQTOQAEIQAENRLEAQSATLIGQOIQTLSSKIVARNESLQQAARSQAQVNTTIG 120  
DB 68 QVSAIQAEQSNLQAEQSLKLEGEITELSKNIVSRNQSLEKQARSQAQVNTTIG 127

QY 121 INAINSKSVDAINRVSAIREVVSANEKMLQOQODKAAVEQKQOENQAANTVAANO 180  
DB 128 INTIVNSKSIETAI SRVAAVMSIVSANNKMLEQKADKKAISEKQVANNDAINTVIANQ 187

QY 181 TTAQNTNALNTQOALEAQLNLQAEELTTAODOKATLVAQKAAAEAEARQAQAAAAEA 240  
DB 188 KLADDAQALTTTQAEKLAELSLAAEKATAEKESKASLLEQKAAAEAEARQAQAAAAEA 246

QY 241 KAAAEAKALQEAQAQAQAANNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNNS 300  
DB 247 ---EKRSQQQSVLASANTNLTAQ-----VQAVSES 274

QY 301 DOESTTATEAOPSASASTAVTANTSSANTYPAGCCTGWKSLAPWVGNVYKNGGQWAA 360  
DB 275 -----AAAPVRKVRPTYSTNASSYPIGECTGWGKTLAPWAGDYWGNGAOWAT 322

QY 361 SAAAAAGYRVGSPSAGAVAVWNDGGYGHVAVVTGQV-GQIQVQEBYVAGNOSIGNYRGW 419  
DB 323 SAAAAAGFTGTPTQVGAIAACWNDGGYGHVAVVTAVESTTRIOVSESNAAGNRTIGNHRGW 382

QY 420 FNP-----GSVSYIY 429  
DB 383 FNPPTTSEGFTYIY 397

RESULT 2

US-09-583-110-4374  
; Sequence 4374, Application US/09583110  
; Patent No. 6699703

; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics  
; FILE REFERENCE: PATH00-07A

; CURRENT APPLICATION NUMBER: US/09/583,110  
; CURRENT FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322

; SEQ ID NO 4374  
; LENGTH: 392  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-4374

Query Match 50.2%; Score 1056.5; DB 2; Length 392;  
Best Local Similarity 50.6%; Pred. No. 1.1e-69;  
Matches 220; Conservative 72; Mismatches 92; Indels 51; Gaps 5;

QY 1 MKKRILSAVLGVLTSSATTLSAVKADDFDAQIASQDSKINNLTAQQAQAQVNTTIG 60  
DB 1 MKKRILASLLSTVWVSQVAVLTTAAHETTTDDKIAAQDNKISNLTAQQAQAQVNTTIG 60

QY 61 QVSALQTOQAEIQAENRLEAQSATLIGQOIQTLSSKIVARNESLQQAARSQAQVNTTIG 120  
DB 68 QVSAIQAEQSNLQAEQSLKLEGEITELSKNIVSRNQSLEKQARSQAQVNTTIG 127

DB 61 QVSAIQAEQSNLQAEQSLKLEGEITELSKNIVSRNQSLEKQARSQAQVNTTIG 120

QY 121 INAINSKSVDAINRVSAIREVVSANEKMLQOQODKAAVEQKQOENQAANTVAANO 180  
DB 128 INTIVNSKSIETAI SRVAAVMSIVSANNKMLEQKADKKAISEKQVANNDAINTVIANQ 187

QY 181 TTAQNTNALNTQOALEAQLNLQAEELTTAODOKATLVAQKAAAEAEARQAQAAAAEA 240  
DB 188 KLADDAQALTTTQAEKLAELSLAAEKATAEKESKASLLEQKAAAEAEARQAQAAAAEA 239

QY 241 KAAAEAKALQEAQAQAQAANNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNNS 300  
DB 247 ---EKRSQQQSVLASANTNLTAQ-----VQAVSES 267

QY 301 DOESTTATEAOPSASASTAVTANTSSANTYPAGCCTGWKSLAPWVGNVYKNGGQWAA 360  
DB 275 -----AAAPVRKVRPTYSTNASSYPIGECTGWGKTLAPWAGDYWGNGAOWAT 315

QY 361 SAAAAAGYRVGSPSAGAVAVWNDGGYGHVAVVTGQV-GQIQVQEBYVAGNOSIGNYRGW 419  
DB 323 SAAAAAGFTGTPTQVGAIAACWNDGGYGHVAVVTAVESTTRIOVSESNAAGNRTIGNHRGW 375

QY 420 FNP-----GSVSYIY 429  
DB 376 FNPPTTSEGFTYIY 390

RESULT 3

US-08-186-222-2  
; Sequence 2, Application US/08186222  
; Patent No. 5559007

; GENERAL INFORMATION:  
; APPLICANT: Suri, Bruno  
; APPLICANT: Schmitz, Albert  
; TITLE OF INVENTION: Bacterial Vectors  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/186,222  
; FILING DATE:

; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/672,205  
; FILING DATE: 19-MAR-1991  
; APPLICATION NUMBER: GB 9006400.7  
; FILING DATE: 22-MAR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Villamizar, JoAnn  
; REGISTRATION NUMBER: 30,598  
; REFERENCE/DOCKET NUMBER: 4-17994/A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (914)785-7121  
; TELEFAX: (914)347-5769

; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 461 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-186-222-2

Query Match 31.9%; Score 671; DB 1; Length 461;



```
Best Local Similarity 35.2%; Pred. No. 2.5e-41;
Matches 169; Conservative 103; Mismatches 140; Indels 68; Gaps 13;

QY 1 MKKRILSAVLVSGVTLSATLISAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60
DB 1 MKKKIISAILMTSVILSAAPLSGVYADT-NSDIAKQDATISSAQSAKAQAQAVDSLOS 59
QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIOTLSSKIVARNESLKQOARSQAQKSNAAATSY 120
DB 60 KVDLSQOKQTSKTAQIAKTESERKALNAQIATILNESIKERTKTLEAQAARSQAQVNSATNY 119
QY 121 INAINSKSVSDAINRVSAIREVWSANEMKLOOEQDQKAAVEQKQOENQAANTVAANOQE 180
DB 120 MDAAVNSKSLTDVIQKVTATITVSSANKQMLEQKEQKELSQKSTVTKKNYNQFVLSQ 179
QY 181 TTAQNTNALNTQOAELEAQLNLQAELETTAAQOKATLVAKAAAEAAEAQAQAAAAA 240
DB 180 SLDQSAQELTSQAELKVATLNYQATIAQDKQALLDEKAAAEKAAQEAQAKQAAYEA 239
QY 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSTEQ-----SAAQ 295
DB 240 Q-----QKEAAQAQAATAAKAVEAATSSASASSAPQVSTSTDTNTTNSASAGN 291
QY 296 AVNSDOESTTATEAOPSASSASTAVVTANTSS-----ANTYP 333
DB 292 SGNSSNSSSSSSSSSSSSSSSNAGNTSGTSGTGTGTTGGSGINSPIGNPYA 351
QY 334 AGOCT---WG-----VKSLAPVGNVNGGOWAAASAAAG--YRVGSTPSAGAVAV 380
DB 352 VGGCTDYVWQYFAAQGIYIRNMP-----GNGGQWASNGPAQGVLVHVGAP--GVIAS 403
QY 381 -----WNDGGYGHVAVYTVQG-QGIQOQAEANYAGNQSGNRYGWFNPGSVSIYPN 431
DB 404 SPSADPVGANGPYGHVAVKSVNSDGTITIKEGGY-GTTWNGHER-TVSASGVTFPMEN 461

RESULT 4
US-09-107-532A-5095
; Sequence 5095, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5095:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 525 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...525
; SEQUENCE DESCRIPTION: SEQ ID NO: 5095:
US-09-107-532A-5095
Query Match 28.2%; Score 594; DB 2; Length 525;
Best Local Similarity 31.0%; Pred. No. 1.3e-35;
Matches 162; Conservative 92; Mismatches 163; Indels 106; Gaps 14;

QY 1 MKKRILSAVLVSGVTLSATLISAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIOG 60
DB 2 VKKSLISAVMVCGSMTLTAVASPIAAADDFDSIQIQDQKADLKNQQAQDAQSIDALES 61
QY 61 QVSALQTOQAELOAENQRLAQSATLGGQIOTLSSKIVARNESLKQOARSQAQKSNAAATSY 120
DB 62 QVSEINTQADLLAKQDTLRQESAQLVKDIADLQERIEKREDTIQKAREAQVSNVTSNY 121
QY 121 INAINSKSVSDAINRVSAIREVWSANEMKLOOEQDQKAAVEQKQOENQAANTVAANOQE 180
DB 122 IDAVLNADSLADAIGRVQAMTMTVKANNNDLMEQKQDKKAVEDKKAENDAKLKEAENQA 181
QY 181 TTAQNTNALNTQOAELEAQLNLQAELETTAAQOKATLVAKA-----AAEE 226
DB 182 ALESQKGDLLSKQADNLVLTSLAAEQATAEDKKADLNKQKAAEAEQAARIREQORLAEQ 241
QY 227 AARQAAAAQAAAEKAAAEKALQEQAAQ-AQAANNNTQATDAS-----DQAAAAADNTQ 281
DB 242 ARQQAQAEKAEKAREQAEAEQAQATQASTAQSSESSATQSSWTESSSATQSSATE 301
QY 282 AAQTGDSTEQSAAQAVNNSDQSTTATEA-----OPSASSASTAVTANTSSANT 331
DB 302 ESTTPESSTESTAPESSESTTAPESSESTTTPESSESTTTPESSESTTTPESSTTEEST 361
QY 332 YPA-----GQCTGWGVKSLAP----- 346
DB 362 TPAPTTPSTDQSDVTGNGT-GSSTPAPTPTTPEQPKVTPAPAPSGSVNGAAIVAEAYK 420
QY 347 WYGN--YMGNG-----GOWAASAAAAAGYRVG-STPSAGAVAV 380
DB 421 YIGTPVWGGKDPGDFDCSGFTRYVMQVTGRDIGGTVPQESAGTKISVSQAKAGDLLF 480
QY 381 W-NDGGYGHVAVYTVQGQIQOQAEANYAGNQSGNRYGWFNFP 422
DB 481 WGSQGGTYHVAIALG--GGQ-YIHAPQPGESVKVGSVQ-WFAP 519

RESULT 5
US-09-222-938A-67
; Sequence 67, Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222,938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 67
; LENGTH: 210
; TYPE: PRT
```





Db 242 KAKQAAKPAKAEVKAEPVASSSTTEAQAAPASSSATESSTQQTETTTPTSDNSATEN 301  
QY 285 TGDSTEQSAQAV-----NNSDQSTTATEAOPSASSASTAVVTANTSSANTYPAGQCT 338  
Db 302 TSSSSEQVPQPTTSDNNGNGQTGGGTVTPTPEPTPAPSADPTINALNLVRSLSIG--- 358  
QY 339 WGVKSLAPVGNWNGGWAASAAAGYRVGSTPSAGAVAVWNGGYGHVAVYVTCVQGG 398  
Db 359 -----LRPV-----WDAGLAASATARAQVEAGGIPNDH-----WSRG--DEVIAIMWAPGN 404  
QY 399 QIQV---QEVYAGNOSIGNYRGW-FNPG 423  
Db 405 SVMWYNETNMVVTASGSG-HRDWEINPG 432  
RESULT 10  
US-09-071-035-484  
; Sequence 484, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB369P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 484:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 422 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-071-035-484  
Query Match 21.8%; Score 458; DB 2; Length 422;  
Best Local Similarity 31.0%; Pred. No. 9.4e-26;  
Matches 131; Conservative 81; Mismatches 166; Indels 44; Gaps 12;  
QY 28 DFDQAQASQDSKINNLTAAQQAQAVNTTIGQVSALQTQQAELQAEQNRLQEAQATLG 87  
Db 1 DNVKKIEKNQESLSLAKAQGLASQVSSLEAEVSVFDESMALEQKQTLKAKSEQIQ 60  
QY 88 QIQTLSSKIVARNESLKOQASAKSNATSVINAINSKSVDAINRVSAIREVVSAN 147  
Db 61 QEITNLNORIEKRNKQADRDVQNGQSTTMDLAVIDADSVADAIKRVQAVSTIVSAN 120  
QY 148 ERMLOQQEQDKAAVEKQKQENQAINTVAANOETIAONTNALNTQQAELAAQLNLAEL 207

Db 121 NDLMOQKEDQAVVDVKAENEKKVKQLQTEAELETKRQDLLLSKSELNVKMSLALQEQ 180  
QY 208 TTAQDKATLVAQKAAAE-EAARQAAAAQAAAE-AKAAAEAKALQEQAAQAAANNNTQ 265  
Db 181 SSAESSKAGLEKQKAAAEQARLAAEQKAAAEKAKQAAAKPAKAEVKAEPVASSSTTE 240  
QY 266 -----ATDASDQAAA-----AADNTQAAQTGDSSTEQSAQAQAV-----NNSDQEST 305  
Db 241 AQAPASSSATESSTQQTETTTPTSDNSATENTGSSSEQVPQPTTSDNNGNGQTGG 300  
QY 306 TATEAOPSASSASTAVVTANTSSANTYPAGQCTGWKSLAPVWNGYNGGOWAASAAAA 365  
Db 301 GTVTPTPEPTPAPSADPTINALNLVRSLSIG-----LRPV-----WDAGLAASATARA 349  
QY 366 GYRVGSTPSAGAVAVWNGGYGHVAVYVTCVQGGIQV---QEVYAGNOSIGNYRGW-FN 421  
Db 350 QVEAGGIPNDH-----WSRG--DEVIAIMWAPGNVIMAWYNETNMVVTASGSG-HRDWEIN 402  
QY 422 PG 423  
Db 403 PG 404  
RESULT 11  
US-10-206-576-484  
; Sequence 484, Application US/10206576  
; Patent No. 6913907  
; GENERAL INFORMATION:  
; APPLICANT: Choi et al.  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 497  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-R  
; COMPUTER: Dell Latitude  
; OPERATING SYSTEM: Windows 98  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/206,576  
; FILING DATE: 29-Jul-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/071,035  
; FILING DATE: 1998-05-04  
; APPLICATION NUMBER: US 60/046,655  
; FILING DATE: 1997-05-16  
; APPLICATION NUMBER: US 60/044,031  
; FILING DATE: 1997-05-06  
; APPLICATION NUMBER: US 60/066,009  
; FILING DATE: 1997-11-14  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hyman, Mark J.  
; REGISTRATION NUMBER: 46,789  
; REFERENCE/DOCKET NUMBER: PB369P1D1  
; INFORMATION FOR SEQ ID NO: 484:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 422 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 484:  
US-10-206-576-484  
Query Match 21.8%; Score 458; DB 2; Length 422;  
Best Local Similarity 31.0%; Pred. No. 9.4e-26;  
Matches 131; Conservative 81; Mismatches 166; Indels 44; Gaps 12;



```
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5035

Query Match      11.2%; Score 235; DB 2; Length 264;
Best Local Similarity 34.5%; Pred. No. 1.2e-09;
Matches 59; Conservative 29; Mismatches 75; Indels 8; Gaps 5;

QY 261 NNNTQATDASDQQAADNTQAAQTGDSFTEQSAQAQVNNSDQESTTATEAOPSASSASTA 320
DB 98 NNNYNNYNNYQ---SNNTSQRTTQPTGGLGASYSTSSNNVHTTSA-PSSNGVSL 153

QY 321 VVTANTSSANTYPAGQCTGCV-KSLAPWGVNNGWGGOWAASAAAAGYRVGSTPSAGAVA 379
DB 154 --NARSASGNLYTSQCCTYYFDRVGGKIGSTWGNANWANAARSGYTVNNSPAKGAIL 211

QY 380 VVNDGGYGHVAVTVGV-QGGQIQOEVANYAGNQSIGNVRGWNPGSVSYIY 429
DB 212 QTSQAGYGHVAYVEGVNNGSIRVSEMYNMGHGVVTSRTISASQAASNY 262

RESULT 15
US-09-134-001C-4539
; Sequence 4539, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4539
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4539
```

```
Query Match      11.2%; Score 235; DB 2; Length 267;
Best Local Similarity 34.5%; Pred. No. 1.2e-09;
Matches 59; Conservative 29; Mismatches 75; Indels 8; Gaps 5;

QY 261 NNNTQATDASDQQAADNTQAAQTGDSFTEQSAQAQVNNSDQESTTATEAOPSASSASTA 320
DB 101 NNNYNNYNNYQ---SNNTSQRTTQPTGGLGASYSTSSNNVHTTSA-PSSNGVSL 156

QY 321 VVTANTSSANTYPAGQCTGCV-KSLAPWGVNNGWGGOWAASAAAAGYRVGSTPSAGAVA 379
DB 157 --NARSASGNLYTSQCCTYYFDRVGGKIGSTWGNANWANAARSGYTVNNSPAKGAIL 214

QY 380 VVNDGGYGHVAVTVGV-QGGQIQOEVANYAGNQSIGNVRGWNPGSVSYIY 429
DB 215 QTSQAGYGHVAYVEGVNNGSIRVSEMYNMGHGVVTSRTISASQAASNY 265
```

Search completed: February 15, 2006, 18:11:59  
Job time : 30.9537 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:29:27 ; Search time 91.8579 Seconds  
(without alignments)  
1960.467 Million cell updates/sec

Title: US-10-797-821-33

Perfect score: 2105

Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNVRGMFPGSVIYPN 431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA\_Main:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*

2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*

3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*

4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*

5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*

6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description          |
|------------|--------|-------------|--------|-------|----------------------|
| 1          | 2105   | 100.0       | 431    | 4     | US-10-383-930-33     |
| 2          | 2105   | 100.0       | 431    | 5     | US-10-797-821-33     |
| 3          | 2096   | 99.6        | 431    | 5     | US-10-383-930-33     |
| 4          | 2096   | 99.6        | 431    | 5     | US-10-797-821-30     |
| 5          | 2092   | 99.4        | 431    | 4     | US-10-383-930-29     |
| 6          | 2092   | 99.4        | 431    | 5     | US-10-797-821-29     |
| 7          | 2075.5 | 98.6        | 432    | 4     | US-10-383-930-31     |
| 8          | 2075.5 | 98.6        | 432    | 5     | US-10-797-821-31     |
| 9          | 2066.5 | 98.2        | 432    | 4     | US-10-383-930-32     |
| 10         | 2066.5 | 98.2        | 432    | 5     | US-10-797-821-32     |
| 11         | 1085.5 | 51.6        | 398    | 5     | US-10-474-792-600    |
| 12         | 1057.5 | 50.2        | 392    | 5     | US-10-472-928-4652   |
| 13         | 1057.5 | 50.2        | 399    | 5     | US-10-617-320-320    |
| 14         | 598    | 28.4        | 524    | 4     | US-10-282-122A-57658 |
| 15         | 562    | 26.7        | 210    | 4     | US-10-154-251-67     |
| 16         | 507    | 24.1        | 449    | 3     | US-09-071-035-482    |
| 17         | 507    | 24.1        | 449    | 4     | US-10-206-576-482    |
| 18         | 507    | 24.1        | 449    | 5     | US-10-912-362-482    |
| 19         | 458    | 21.8        | 422    | 3     | US-09-071-035-484    |
| 20         | 458    | 21.8        | 422    | 4     | US-10-206-576-484    |
| 21         | 458    | 21.8        | 422    | 5     | US-10-912-362-484    |
| 22         | 335.5  | 15.9        | 630    | 5     | US-10-494-674-6      |
| 23         | 329.5  | 15.7        | 600    | 3     | US-09-738-626-5197   |
| 24         | 301    | 14.3        | 609    | 5     | US-10-501-282-3184   |
| 25         | 300    | 14.3        | 440    | 5     | US-10-501-282-3056   |
| 26         | 297.5  | 14.1        | 422    | 5     | US-10-501-282-3054   |
| 27         | 288.5  | 13.7        | 377    | 5     | US-10-501-282-3052   |

ALIGNMENTS

RESULT 1

US-10-383-930-33

; Sequence 33, Application US/10383930

; Publication No. US20040127400A1

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J

; APPLICANT: Taubman, Martin A

; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein

; FILE REFERENCE: 25669-018

; CURRENT APPLICATION NUMBER: US/10/383,930

; CURRENT FILING DATE: 2003-03-07

; PRIOR APPLICATION NUMBER: 60/402,483

; PRIOR FILING DATE: 2002-08-08

; PRIOR APPLICATION NUMBER: 60/363,209

; PRIOR FILING DATE: 2002-03-07

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 33

; LENGTH: 431

; TYPE: PRT

; ORGANISM: Streptococcus mutans

; US-10-383-930-33

Query Match 100.0%; Score 2105; DB 4; Length 431;

Best Local Similarity 100.0%; Pred. No. 2e-120;

Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKKRILSAVLVSGVTLSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG 60

Db 1 MKKRILSAVLVSGVTLSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG 60

Qy 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIIVARNESLKQARSAQKSNATSY 120

Db 61 QVSALQTOQAELOAENORLEAQSATLGGQIQTLSSKIIVARNESLKQARSAQKSNATSY 120

Qy 121 INAIINSSVSDAINRVSAIREVVSANEKMLQQQODKAAVVEQKQENQAAINTVAANQE 180

Db 121 INAIINSSVSDAINRVSAIREVVSANEKMLQQQODKAAVVEQKQENQAAINTVAANQE 180

Qy 181 TTAQNTNALTQQAQLEAAQLNLQALTTAOPDKATLVAQKAAAEFAAFQAAAAQAAAEA 240

Db 181 TTAQNTNALTQQAQLEAAQLNLQALTTAOPDKATLVAQKAAAEFAAFQAAAAQAAAEA 240

Qy 241 KAAAEKALQEQAQAQAQAANNTQATDASDQAAAAADNTQAAQTGDSQAAQAVNNS 300

Db 241 KAAAEKALQEQAQAQAQAANNTQATDASDQAAAAADNTQAAQTGDSQAAQAVNNS 300

Qy 301 DOESTTATEAOPSASSASTAVVTANTSSANTYPAGCTGWGKSLAPWVGNNGSGOWAA 360

Db 301 DOESTTATEAOPSASSASTAVVTANTSSANTYPAGCTGWGKSLAPWVGNNGSGOWAA 360

QY 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTVGGGQIOVQEANYAGNOSIGNYRGWF 420  
DB 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTVGGGQIOVQEANYAGNOSIGNYRGWF 420  
QY 421 NPGSVSYIYPN 431  
DB 421 NPGSVSYIYPN 431

RESULT 2  
US-10-797-821-33  
; Sequence 33, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 33  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-821-33

Query Match 100.0%; Score 2105; DB 5; Length 431;  
Best Local Similarity 100.0%; Pred. No. 2e-120;  
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTTIG 60  
DB 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTTIG 60  
QY 61 QVSALQTQOAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQVNTTIG 120  
DB 61 QVSALQTQOAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQVNTTIG 120  
QY 121 INAIINSKVSDAINRVSAIREVWSANEKMLQQEQDKAAVEQKQENQAANTVAANQE 180  
DB 121 INAIINSKVSDAINRVSAIREVWSANEKMLQQEQDKAAVEQKQENQAANTVAANQE 180  
QY 181 TTAQNTNALNTQOAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQVNTTIG 240  
DB 181 TTAQNTNALNTQOAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQVNTTIG 240  
QY 241 KAAAEKALQOAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQVNTTIG 300  
DB 241 KAAAEKALQOAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQVNTTIG 300  
QY 301 DQESTTATEAQSASSASTAVTANTSSANTYPAGCTWGVKSLAPWVGNVWNGGQWAA 360  
DB 301 DQESTTATEAQSASSASTAVTANTSSANTYPAGCTWGVKSLAPWVGNVWNGGQWAA 360  
QY 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTVGGGQIOVQEANYAGNOSIGNYRGWF 420  
DB 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTVGGGQIOVQEANYAGNOSIGNYRGWF 420

QY 421 NPGSVSYIYPN 431  
DB 421 NPGSVSYIYPN 431  
RESULT 3  
US-10-383-930-30  
; Sequence 30, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 30  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-30

Query Match 99.6%; Score 2096; DB 4; Length 431;  
Best Local Similarity 99.5%; Pred. No. 7.1e-120;  
Matches 429; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTTIG 60  
DB 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTTIG 60  
QY 61 QVSALQTQOAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQVNTTIG 120  
DB 61 QVSALQTQOAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQVNTTIG 120  
QY 121 INAIINSKVSDAINRVSAIREVWSANEKMLQQEQDKAAVEQKQENQAANTVAANQE 180  
DB 121 INAIINSKVSDAINRVSAIREVWSANEKMLQQEQDKAAVEQKQENQAANTVAANQE 180  
QY 181 TTAQNTNALNTQOAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQVNTTIG 240  
DB 181 TTAQNTNALNTQOAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQVNTTIG 240  
QY 241 KAAAEKALQOAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQVNTTIG 300  
DB 241 KAAAEKALQOAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSQAQVNTTIG 300  
QY 301 DQESTTATEAQSASSASTAVTANTSSANTYPAGCTWGVKSLAPWVGNVWNGGQWAA 360  
DB 301 DQESTTATEAQSASSASTAVTANTSSANTYPAGCTWGVKSLAPWVGNVWNGGQWAA 360  
QY 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTVGGGQIOVQEANYAGNOSIGNYRGWF 420  
DB 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAVYVTVGGGQIOVQEANYAGNOSIGNYRGWF 420  
QY 421 NPGSVSYIYPN 431  
DB 421 NPGSVSYIYPN 431  
RESULT 4  
US-10-797-821-30  
; Sequence 30, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens



```

; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; PRIOR FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-30

      Query Match          99.6%; Score 2096; DB 5; Length 431;
      Best Local Similarity 99.5%; Pred. No. 7.1e-120;
      Matches 429; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60
DB 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQOARSQKSNATSY 120
DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQOARSQKSNATSY 120

QY 121 INAIINSKVSDAINRVSAREVWSANEKMLQQOQODKAAVEQKQOENQAINTVAANOE 180
DB 121 INAIINSKVSDAINRVSAREVWSANEKMLQQOQODKAAVEQKQOENQAINTVAANOE 180

QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEITTAQDOKATLVAQKAAAEAAQAAAAQA 240
DB 181 TTAQNTNALNTQQAQLEAAQLNLQAEITTAQDOKATLVAQKAAAEAAQAAAAQA 240

QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQA 300
DB 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQA 300

QY 301 DOESTTATEAOPSSASASTAVVTANTSSANTYPAGCTWGVKSLAPWVGNVWNGGOWAA 360
DB 301 DOESTTATEAOPSSASASTAVVTANTSSANTYPAGCTWGVKSLAPWVGNVWNGGOWAA 360

QY 361 SAAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVQEAANYAGNQSIGNYRGWF 420
DB 361 SAAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVQEAANYAGNQSIGNYRGWF 420

QY 421 NPGSVSYIYPN 431
DB 421 NPGSVSYIYPN 431

RESULT 5
US-10-383-930-29
; Sequence 29, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-29

      Query Match          99.4%; Score 2092; DB 4; Length 431;
      Best Local Similarity 99.3%; Pred. No. 1.3e-119;
      Matches 428; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60
DB 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQOARSQKSNATSY 120
DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIQITLSSKIVARNESLKQOARSQKSNATSY 120

QY 121 INAIINSKVSDAINRVSAREVWSANEKMLQQOQODKAAVEQKQOENQAINTVAANOE 180
DB 121 INAIINSKVSDAINRVSAREVWSANEKMLQQOQODKAAVEQKQOENQAINTVAANOE 180

QY 181 TTAQNTNALNTQQAQLEAAQLNLQAEITTAQDOKATLVAQKAAAEAAQAAAAQA 240
DB 181 TTAQNTNALNTQQAQLEAAQLNLQAEITTAQDOKATLVAQKAAAEAAQAAAAQA 240

QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQA 300
DB 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGDSQSAQA 300

QY 301 DOESTTATEAOPSSASASTAVVTANTSSANTYPAGCTWGVKSLAPWVGNVWNGGOWAA 360
DB 301 DOESTTATEAOPSSASASTAVVTANTSSANTYPAGCTWGVKSLAPWVGNVWNGGOWAA 360

QY 361 SAAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVQEAANYAGNQSIGNYRGWF 420
DB 361 SAAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVQEAANYAGNQSIGNYRGWF 420

QY 421 NPGSVSYIYPN 431
DB 421 NPGSVSYIYPN 431

RESULT 6
US-10-797-821-29
; Sequence 29, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
```

; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-821-29

Query Match  
Best Local Similarity 99.4%; Score 2092; DB 5; Length 431;  
Matches 428; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTIQG 60  
DB 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTIQG 60

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSNAAATSY 120  
DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOQODKAAVEQKQOENQAAINTVAANQE 180  
DB 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOQODKAAVEQKQOENQAAINTVAANQE 180

QY 181 TIAQNTNALNTQQAQLEAAQNLQAEELTTAODQKATLVAQKAAABEAAARQAAAAQAAREA 240  
DB 181 TIAQNTNALNTQQAQLEAAQNLQAEELTTAODQKATLVAQKAAABEAAARQAAAAQAAREA 240

QY 241 KAAAEKALQEQAAQAAQAAA-NNNTQATDASDQOAAAADNTQAAOTGDSQSAQAQAVNN 299  
DB 241 KAAAEKALQEQAAQAAQAAA-NNNTQATDASDQOAAAADNTQAAOTGDSQSAQAQAVNN 300

QY 300 SDQESTTATEAQPSSASTAVTANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGOWA 359  
DB 300 SDQESTTATEAQPSSASTAVTANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGOWA 360

QY 360 ASAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVQEAANYAGNQSIGNYRGW 419  
DB 360 ASAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVQEAANYAGNQSIGNYRGW 420

QY 420 NPGSVSVIYPN 431  
DB 421 NPGSVSVIYPN 432

RESULT 7  
US-10-383-930-31  
; Sequence 31, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 31  
; LENGTH: 432  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-31

Query Match  
Best Local Similarity 98.6%; Score 2075.5; DB 4; Length 432;  
Matches 426; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTIQG 60  
DB 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTIQG 60

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSNAAATSY 120  
DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOQODKAAVEQKQOENQAAINTVAANQE 180  
DB 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOQODKAAVEQKQOENQAAINTVAANQE 180

QY 181 TIAQNTNALNTQQAQLEAAQNLQAEELTTAODQKATLVAQKAAABEAAARQAAAAQAAREA 240  
DB 181 TIAQNTNALNTQQAQLEAAQNLQAEELTTAODQKATLVAQKAAABEAAARQAAAAQAAREA 240

QY 241 KAAAEKALQEQAAQAAQAAA-NNNTQATDASDQOAAAADNTQAAOTGDSQSAQAQAVNN 299  
DB 241 KAAAEKALQEQAAQAAQAAA-NNNTQATDASDQOAAAADNTQAAOTGDSQSAQAQAVNN 300

QY 300 SDQESTTATEAQPSSASTAVTANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGOWA 359  
DB 300 SDQESTTATEAQPSSASTAVTANTSSANTYPAGQCTGWGKSLAPWVGNWNGGGOWA 360

QY 360 ASAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVQEAANYAGNQSIGNYRGW 419  
DB 360 ASAAAAGYRVGSTPSAGAVAVWMDGGYGHVAVYVTVGGGQIQVQEAANYAGNQSIGNYRGW 420

QY 420 NPGSVSVIYPN 431  
DB 421 NPGSVSVIYPN 432

RESULT 8  
US-10-797-821-31  
; Sequence 31, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 31  
; LENGTH: 432  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-821-31

Query Match  
Best Local Similarity 98.6%; Score 2075.5; DB 5; Length 432;  
Matches 426; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTIQG 60  
DB 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTIQG 60

QY 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSNAAATSY 120  
DB 61 QVSALQTOQAELOAENORLEAQSATLGGQIOTLSSKIVARNESLKQOARSAQKSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOQODKAAVEQKQOENQAAINTVAANQE 180

Db 121 INAIINSKSVDAINRVSAIREVVSANEKMLQOQEQDKAAVEKQOENQAAINTVAANQE 180  
Qy 181 TTAQNTNALNTQOAEAAQALNLQAELETTTAQDQKATLVAKKAAAEBAARQAAAAQAAAA 240  
Db 181 TTAQNTNALNTQOAEAAQALNLQAELETTTAQDQKATLVAKKAAAEBAARQAAAAQAAAA 240  
Qy 241 KAAAEAKALQEQAQAQAAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299  
Db 241 KAAAEAKALQEQAQAQAAAAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 300  
Qy 300 SQESTTATEAOPSSASSASTAVVTANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 359  
Db 301 SQESTTATAOPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 360  
Qy 360 ASAAAAGRVGSTPSAGAVVWNDGGYGHVAVVTGVGGQIQVQEANYAGNOSIGNYRGW 419  
Db 361 ASAAAAGRVGSTPSAGAVVWNDGGYGHVAVVTGVGGQIQVQEANYAGNOSIGNYRGW 420  
Qy 420 FNPGSVSIYPN 431  
Db 421 FNPGSVSIYPN 432  
RESULT 9  
US-10-383-930-32  
; Sequence 32, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 32  
; LENGTH: 432  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-32  
Query Match 98.2%; Score 2066.5; DB 4; Length 432;  
Best Local Similarity 98.4%; Pred. No. 4.5e-118;  
Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;  
Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQAVNTIOG 60  
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQAVNTIOG 60  
Qy 61 QVSALQTQOAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120  
Db 61 QVSALQTQOAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120  
Qy 121 INAIINSKSVDAINRVSAIREVVSANEKMLQOQEQDKAAVEKQOENQAAINTVAANQE 180  
Db 121 INAIINSKSVDAINRVSAIREVVSANEKMLHQQEQDKAAVEKQOENQAAINTVAANQE 180  
Qy 181 TTAQNTNALNTQOAEAAQALNLQAELETTTAQDQKATLVAKKAAAEBAARQAAAAQAAAA 240  
Db 181 TTAQNTNALNTQOAEAAQALNLQAELETTTAQDQKATLVAKKAAAEBAARQAAAAQAAAA 240  
Qy 241 KAAAEAKALQEQAQAQAAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299  
Db 241 KAAAEAKALQEQAQAQAAAAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 300  
Qy 300 SQESTTATEAOPSSASSASTAVVTANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 359  
Db 301 SQESTTATAOPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 360  
Qy 360 ASAAAAGRVGSTPSAGAVVWNDGGYGHVAVVTGVGGQIQVQEANYAGNOSIGNYRGW 419  
Db 361 ASAAAAGRVGSTPSAGAVVWNDGGYGHVAVVTGVGGQIQVQEANYAGNOSIGNYRGW 420

Db 301 SQESTTATAOPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 360  
Qy 360 ASAAAAGRVGSTPSAGAVVWNDGGYGHVAVVTGVGGQIQVQEANYAGNOSIGNYRGW 419  
Db 361 ASAAAAGRVGSTPSAGAVVWNDGGYGHVAVVTGVGGQIQVQEANYAGNOSIGNYRGW 420  
Qy 420 FNPGSVSIYPN 431  
Db 421 FNPGSVSIYPN 432  
RESULT 10  
US-10-797-821-32  
; Sequence 32, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 32  
; LENGTH: 432  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-821-32  
Query Match 98.2%; Score 2066.5; DB 5; Length 432;  
Best Local Similarity 98.4%; Pred. No. 4.5e-118;  
Matches 425; Conservative 1; Mismatches 5; Indels 1; Gaps 1;  
Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQAVNTIOG 60  
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQAVNTIOG 60  
Qy 61 QVSALQTQOAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120  
Db 61 QVSALQTQOAELOAENORLEAQSATLGGQIQTLSSKIVARNESLKQOARSAQKSNAAATSY 120  
Qy 121 INAIINSKSVDAINRVSAIREVVSANEKMLQOQEQDKAAVEKQOENQAAINTVAANQE 180  
Db 121 INAIINSKSVDAINRVSAIREVVSANEKMLHQQEQDKAAVEKQOENQAAINTVAANQE 180  
Qy 181 TTAQNTNALNTQOAEAAQALNLQAELETTTAQDQKATLVAKKAAAEBAARQAAAAQAAAA 240  
Db 181 TTAQNTNALNTQOAEAAQALNLQAELETTTAQDQKATLVAKKAAAEBAARQAAAAQAAAA 240  
Qy 241 KAAAEAKALQEQAQAQAAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299  
Db 241 KAAAEAKALQEQAQAQAAAAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 300  
Qy 300 SQESTTATEAOPSSASSASTAVVTANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 359  
Db 301 SQESTTATAOPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 360  
Qy 360 ASAAAAGRVGSTPSAGAVVWNDGGYGHVAVVTGVGGQIQVQEANYAGNOSIGNYRGW 419  
Db 361 ASAAAAGRVGSTPSAGAVVWNDGGYGHVAVVTGVGGQIQVQEANYAGNOSIGNYRGW 420







GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 18:30:27 ; Search time 8.32046 Seconds  
(without alignments)  
736.166 Million cell updates/sec

Title: US-10-797-821-33

Perfect score: 2105

Sequence: 1 MKKRILSAVLVSGVTLSSAT.....SIGNRGMFNGSVIYPN 431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 107799 seqs, 14211699 residues

Total number of hits satisfying chosen parameters: 107799

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_New.\*

- 1: /cgn2\_6/prodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/prodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/prodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/prodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 5: /cgn2\_6/prodata/2/pubpaa/US05\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/prodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 7: /cgn2\_6/prodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/prodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description                           |
|------------|--------|-------------|--------|-------|---------------------------------------|
| 1          | 2092   | 99.4        | 431    | 7     | US-11-052-554A-210 Sequence 210, App  |
| 2          | 1092.5 | 51.9        | 398    | 7     | US-11-052-554A-252 Sequence 252, App  |
| 3          | 436    | 20.7        | 211    | 7     | US-11-052-554A-352 Sequence 352, App  |
| 4          | 413.5  | 19.6        | 544    | 7     | US-11-052-554A-358 Sequence 358, App  |
| 5          | 235    | 11.2        | 257    | 6     | US-10-793-626-3244 Sequence 3244, App |
| 6          | 227    | 10.8        | 971    | 7     | US-11-052-554A-3 Sequence 3, Appl     |
| 7          | 226    | 10.7        | 394    | 7     | US-11-052-554A-79 Sequence 79, Appl   |
| 8          | 225.5  | 10.7        | 1236   | 6     | US-10-873-528-109 Sequence 109, App   |
| 9          | 215    | 10.2        | 149    | 6     | US-10-793-626-1682 Sequence 1682, App |
| 10         | 215    | 10.2        | 157    | 6     | US-10-793-626-2870 Sequence 2870, App |
| 11         | 215    | 10.2        | 270    | 6     | US-10-485-517-413 Sequence 413, App   |
| 12         | 205    | 9.7         | 1562   | 7     | US-11-052-554A-211 Sequence 211, App  |
| 13         | 198.5  | 9.4         | 2101   | 6     | US-10-857-780-23 Sequence 23, Appl    |
| 14         | 196.5  | 9.3         | 1095   | 6     | US-10-793-626-3154 Sequence 3154, App |
| 15         | 192.5  | 9.1         | 1448   | 6     | US-10-485-517-212 Sequence 212, App   |
| 16         | 189    | 9.0         | 1107   | 6     | US-10-485-517-145 Sequence 145, App   |
| 17         | 187    | 8.9         | 758    | 6     | US-10-485-517-144 Sequence 144, App   |
| 18         | 184    | 8.7         | 1586   | 6     | US-10-821-234-901 Sequence 901, App   |
| 19         | 180    | 8.6         | 1410   | 6     | US-10-878-556A-136 Sequence 136, App  |
| 20         | 180    | 8.6         | 1122   | 6     | US-10-467-657-6112 Sequence 6112, App |
| 21         | 179.5  | 8.5         | 330    | 6     | US-10-485-517-415 Sequence 415, App   |
| 22         | 176.5  | 8.4         | 1126   | 6     | US-10-485-517-248 Sequence 248, App   |
| 23         | 176    | 8.4         | 3712   | 7     | US-11-019-711-48 Sequence 48, Appl    |
| 24         | 176    | 8.4         | 3712   | 7     | US-11-019-711-51 Sequence 51, Appl    |
| 25         | 173    | 8.2         | 5024   | 6     | US-10-793-626-2964 Sequence 2964, App |

ALIGNMENTS

RESULT 1

US-11-052-554A-210  
; Sequence 210, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; PRIOR FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 210  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans UA159  
US-11-052-554A-210

|                       |       |              |          |            |     |        |     |
|-----------------------|-------|--------------|----------|------------|-----|--------|-----|
| Query Match           | 99.4% | Score        | 2092     | DB         | 7   | Length | 431 |
| Best Local Similarity | 99.3% | Pred. No.    | 7.6e-113 |            |     |        |     |
| Matches               | 428   | Conservative | 0        | Mismatches | 3   | Indels | 0   |
| Gaps                  | 0     |              |          |            |     |        |     |
| QY                    | 1     | 1            | 1        | 1          | 1   | 1      | 1   |
| DB                    | 1     | 1            | 1        | 1          | 1   | 1      | 1   |
| QY                    | 61    | 61           | 61       | 61         | 61  | 61     | 61  |
| DB                    | 61    | 61           | 61       | 61         | 61  | 61     | 61  |
| QY                    | 121   | 121          | 121      | 121        | 121 | 121    | 121 |
| DB                    | 121   | 121          | 121      | 121        | 121 | 121    | 121 |
| QY                    | 181   | 181          | 181      | 181        | 181 | 181    | 181 |
| DB                    | 181   | 181          | 181      | 181        | 181 | 181    | 181 |
| QY                    | 241   | 241          | 241      | 241        | 241 | 241    | 241 |
| DB                    | 241   | 241          | 241      | 241        | 241 | 241    | 241 |
| QY                    | 301   | 301          | 301      | 301        | 301 | 301    | 301 |
| DB                    | 301   | 301          | 301      | 301        | 301 | 301    | 301 |

Sequence 141, App  
Sequence 58, Appl  
Sequence 56, Appl  
Sequence 2704, App  
Sequence 252, App  
Sequence 212, App  
Sequence 59, Appl  
Sequence 146, App  
Sequence 2098, App  
Sequence 3988, App  
Sequence 1052, App  
Sequence 1400, App  
Sequence 223, App  
Sequence 8, Appl  
Sequence 1016, App  
Sequence 1015, App  
Sequence 1076, App  
Sequence 231, App  
Sequence 20, Appl

Db 301 DQESTTATEAQSASASTAAVAANTSSANTYPAGCTGKSLAPWGVNNGGQWAA 360  
|||||

QY 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAYVTGQGGIQVQOEANYAGNQSIGNYRGWF 420  
|||||

Db 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAYVTGQGGIQVQOEANYAGNQSIGNYRGWF 420  
|||||

QY 421 NPGSVSYIYPN 431  
|||||

Db 421 NPGSVSYIYPN 431  
|||||

## RESULT 2

US-11-052-554A-252

; Sequence 252, Application US/11052554A

; Publication No. US20050288866A1

; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

; FILE REFERENCE: 30853/40359A

; CURRENT APPLICATION NUMBER: US/11/052,554A

; PRIOR FILING DATE: 2005-02-07

; PRIOR APPLICATION NUMBER: US 60/589,227

; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; PRIOR FILING DATE: 2004-02-06

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 252

; LENGTH: 398

; TYPE: PRT

; ORGANISM: Streptococcus pyogenes MGAS8232

US-11-052-554A-252

Query Match 51.9%; Score 1092.5; DB 7; Length 398;

Best Local Similarity 53.7%; Pred. No. 7e-56;

Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTTQG 60  
|||||Db 1 MKKRILSAVLVSGVTILGAATT---VCAEDLSTKIAKQSIISNLTAEQKAAQNOVSALQA 57  
|||||QY 61 QVSALQTQAEIQAENORLEAQSATILGQOIOTLSSKIVARNESLAKOARSACKSNAATSY 120  
|||||Db 58 QVSSIQSEQDKUTARNTLEALSKEPQEBIKALTSQIVARNEKLNQAKSYKNNETSGY 117  
|||||QY 121 INAINSKSVDAINRVSAIREVWSANEKWLQQEQDKAAVEQKQOENQAAINTVAANOE 180  
|||||Db 118 INALLNSKISDVNRLVAINRAVSANAKLLLEQKADKVSLEKQAAQTALNTIAANWA 177  
|||||QY 181 TTAQNTNALNTQOALEAAQLNLQAEITPAQDKATLVAQKAAABEAAKQAAAAQAAAEA 240  
|||||Db 178 MAEENQTLTQANLEAATANLALQASATEDKANLVAQKAAEAKAAEALAQQAQKV 237  
|||||QY 241 KAAAEKALQEOAAQAAQAAANNNTQATDASDQOAAAADNTQAAOTGDSSTEQSAAQVNN 300  
|||||Db 238 KAQEQA---AQQAASVEAAKSAITPAQA-----YDSNTYPVQCCTGAKSLAPWAGNNGGQWAA 280  
|||||QY 301 DQESTTATEAQSASASTAVVTVANTSSANTYPAGCTGKSLAPWGVNNGGQWAA 360  
|||||Db 281 ----PAASARQTS-----YDSNTYPVQCCTGAKSLAPWAGNNGGQWAA 326  
|||||QY 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAYVTGQGGIQVQOEANYAGNQSIGNYRGW 419  
|||||Db 327 SAQAAAGYRTGTPMVGAIVAVNDGGYGHVAVVVEVQSASSIRVMESNSGRQYIADHRG 386  
|||||QY 420 NPGSVSYIYPN 431  
|||||Db 387 FNPTGVTFIYPH 398  
|||||

## RESULT 3

US-11-052-554A-352

; Sequence 352, Application US/11052554A

; Publication No. US20050288866A1

; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

; FILE REFERENCE: 30853/40359A

; CURRENT APPLICATION NUMBER: US/11/052,554A

; CURRENT FILING DATE: 2005-02-07

; PRIOR APPLICATION NUMBER: US 60/589,227

; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; PRIOR FILING DATE: 2004-02-06

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 352

; LENGTH: 211

; TYPE: PRT

; ORGANISM: Streptococcus mutans UA159

US-11-052-554A-352

Query Match 20.7%; Score 436; DB 7; Length 211;

Best Local Similarity 44.4%; Pred. No. 9.5e-19;

Matches 95; Conservative 28; Mismatches 57; Indels 34; Gaps 7;

QY 247 KALQEOAAQAAQAAANNNTQATDASDQOAAAADNTQAAQOTGDSSTEQSAAQ-----AV 297  
|||||Db 3 KQFLEKAVFTVAAA-----TAATVVLGNKQADAD-TYTLQEGDSFSSVAQRYHMDAYELASM 57  
|||||QY 298 NNSDOES-----TTATEAQPSSASASTAVVT-ANTSSANTYPAGCTGKVKSL 344  
|||||Db 58 NGKDTISLILPGQTLTVNGSRAAPDNQAAAPTDTTQATTETNDANANTYPVGCTGKVKAV 117  
|||||QY 345 APWGVNNGGQWAAASAAAAGYRVGSTPSAGAVAVNDGGYGHVAYVTGV-OGGOIQVQ 403  
|||||Db 118 ATWAGDMWNGGDWASSASAOQYTVGNTPAVGSIMCMTDGGYGHVAYVTAVGEDGKVQL 177  
|||||QY 404 EANYAGNQSIGNYRGWFN-----PGSVSYIYPN 431  
|||||Db 178 ESNYKQDQWVDNRYRGWFDPNNSGTGPGSVSYIYPN 211  
|||||

## RESULT 4

US-11-052-554A-358

; Sequence 358, Application US/11052554A

; Publication No. US20050288866A1

; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

; FILE REFERENCE: 30853/40359A

; CURRENT APPLICATION NUMBER: US/11/052,554A

; CURRENT FILING DATE: 2005-02-07

; PRIOR APPLICATION NUMBER: US 60/589,227

; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 358

; LENGTH: 544

; TYPE: PRT

; ORGANISM: Streptococcus mutans UA159

US-11-052-554A-358

Query Match 19.6%; Score 413.5; DB 7; Length 544;

Best Local Similarity 25.4%; Pred. No. 5.4e-17;

Matches 143; Conservative 74; Mismatches 192; Indels 153; Gaps 17;

QY 4 RILSAVLVSGVTLSATTLSAVKADDFDAQIA-----SODSKINNLTAAQOAAQAVNTIQ 59  
|||||Db 2 RKLKVALFASILGLMLAVSSYTAADTEDNQVTISHYNEQAGTFDYNVQAANGKTIQSID 61  
|||||



```

QY      60  GOVSALOTQQAELQ-----AENORLEAQ-----ATL 86
      : : : :
Db      62  VAIWSEENGQDLKWYHASNSQLTVHFNAENHSGKVGSYIAHAYITVTDGNRVGNL 121
      : : : :
QY      87  GOOIOTLASKIVARNESLKQ---QARSAQSNAAITSYINAIINSKVSDAINRVA--- 139
      : : : :
Db     122  GKRLKLSL-----APQLSKQGLQLFQSKLKPSADQLFSAVWDENGQDLDLHWYTADAG 177
      : : : :
QY     140  -----IREVVSANEKMLQOQEQD---KAAVE-QKQEQENQAIAINTVAAN 178
      : : : :
Db     178  NTLAGVANHKGYGTVHVHTYLKQNGKMIPIPSAQDIDIPKPKVIQDKINDTSDYDVVVNN 237
      : : : :
QY     179  -----QETTAQNTNALNTQQAQLEAAQLNQAE--- 206
      : : : :
Db     238  VPPYISSVAIPWSENGQDLKWYQATKVADGIFKTVYLVKTHRFELGNYQAHYGDQ 297
      : : : :
QY     207  -----LTTAQDQKATL---VAQKAAAAEAAQAAAAAQAAAAEAKAAAA 246
      : : : :
Db     298  LSKLDLGGETHFNVPSIINVEDPQVTDIHNINKGTFDVTVAETDNSKAIQISIAWVS 357
      : : : :
QY     247  KALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTG-----DSTEQSAAQ 295
      : : : :
Db     358  DANQANLYWEAKQLANGAKAII TVDQ-----KHGNQTGSYNVHVYHNDGTTSGHVL 411
      : : : :
QY     296  AVNNSDQESTTATEAOPASASASTAVTANTSSANTYPAGCOTWGVKSLAPWGVNWXNG 355
      : : : :
Db     412  A-----NOQLNQTVHYQPSA-----VRITAYMNEKNTYPVGQCOTWGVKELAPWIPNLGNG 462
      : : : :
QY     356  GOWASAAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYTVGQ-GGOIQVOQEAANYAGNQSTG 414
      : : : :
Db     463  GOWASTVAVGFKGTGTPVKVGAICWSDGGYGHVAVYTHVESNNRLQVKEANYKQOYIS 522
      : : : :
QY     415  NTRGWFPN-----GSVSYIYPN 431
      : : : :
Db     523  NFRGWFDPDTTSYLGRLTVIYPD 544
      : : : :

```

## RESULT 5

```

US-10-793-626-3244
; Sequence 3244, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS3480US
; CURRENT APPLICATION NUMBER: US/10793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3244
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-3244

```

|                       |                  |                    |           |             |
|-----------------------|------------------|--------------------|-----------|-------------|
| Query Match           | 11.2%            | Score 235;         | DB 6;     | Length 257; |
| Best Local Similarity | 34.5%;           | Pred. No. 3.5e-07; |           |             |
| Matches 59;           | Conservative 29; | Mismatches 75;     | Indels 8; | Gaps 5;     |

|  |    |    |     |  |     |
|--|----|----|-----|--|-----|
|  | Qy | Db | 261 | NNNTQTADSOOAAAANDNTOAAGCTDSTEGSAAAVNNSDQSRTTATTAQAQPSASSASTA   | 320 |
|  |    |    |     | : : : : : : :  | :   |
|  |    |    |     | : : : : : : :  | :   |
|  |    |    |     | : : : : ~ : :  | :   |
|  |    |    | 91  | NNYSNYNNYNVQQ--SSNSTQRKTPTGGLGASYSTSSNVHVTTSTA-PESNGVSUS       | 146 |
|  |    |    |     | : : : : : : :  | :   |
|  |    |    |     | : : : : : : :  | :   |
|  |    |    |     | : : : : ~ : :  | :   |
|  |    |    | 321 | VVTANTSSANTYPACQGTWVG-KSLAPVMGVNWGNQGWAASAAAAAGRYGSTPSPSAGA VA | 379 |
|  |    |    |     | : : : : : : :  | :   |
|  |    |    |     | : : : : ~ : :  | :   |
|  |    |    | 147 | --NARSAGNLYSQCITYFYDFRGGVKIGSTGWGANNNNAANAARGSYTVNNSPKAGAIL    | 204 |
|  |    |    |     | : : : : : : :  | :   |
|  |    |    |     | : : : : ~ : :  | :   |
|  |    |    |     | : : : : ~ : :  | :   |

[illegible]

|                           |       |                    |          |             |
|---------------------------|-------|--------------------|----------|-------------|
| Query Match               | 10.8% | Score 227;         | DB 7;    | Length 971; |
| Best Local Similarity     | 23.8% | Pred. No. 4.5e-06; |          |             |
| Matches 105; Conservative | 76;   | Mismatches 205;    |          |             |
|                           |       | Indels 56;         | Gaps 10; |             |

|  |    |     |   |     |
|--|----|-----|---|-----|
|  | QY | 19  | ATTLSAVKADDFDAIQIASODSKINNLLTAQQOQAQAQVNTTGGVSALQTQOAELQAEHQNR  | 79  |
|  | DB | 121 | AQNTAAAKKSASDASTSAREAAHTATDAADSARAASTSAGQAASSAQSSASGATSTTKA     | 180 |
|  | QY | 79  | LEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAOKSNAATSYINAIINSKSVSDAINRVS | 138 |
|  | DB | 181 | TFA-----SKSAABESSKAATSAGA-AKTSETNAAVSOQSAA-----TS               | 221 |
|  | QY | 139 | AIREVVSANERKMLOQEQBDKAAVEKQKOENQA--IN-TVAANOETIAQN-----TNA      | 188 |
|  | DB | 222 | ASTATTTKASEAARGDASASKEAAKSGSETSAASSASAASSATAAGNSAKAAKTSETNA     | 281 |
|  | QY | 189 | LNTQQAQLE-----AAQLNLQALELTITTAQDDOKATLVQAQAAAEEAARQAQAAA- - - - | 234 |
|  | DB | 282 | KSEETAEDQSSAAGAAGSKTTAALLSASASTSAGQASASATPAGKSAESAASSASTATTKA   | 341 |
|  | QY | 235 | -----QAAAAEAKAEAKALQEQAQAQAANNNTQTATDASDOQAAAAADNTQAOTGDST      | 289 |
|  | DB | 342 | GEATEQAASAAASASAAKTSETNKASETSAESSKTTAAASASSASAASSASSASASKDEAT   | 401 |
|  | QY | 290 | EQSAQAQVNNSDQESTTATEQAQPSASASASTAVVTANTSSANTYPACOGCTGWVKSLAPWVG | 349 |
|  | DB | 402 | RQASA-AKSSWATTATKATEAAGSATAAQSKSTAESAATRAETAakraEDIAASVALED     | 460 |
|  | QY | 350 | NYWNGG--QWAASAAAAGYRVGSTPSAGAVAVMNDGGYGHVAYVTGVCGGQITQVOEANY    | 407 |
|  | DB | 461 | ASTTKKGIVQLSSATNSTSESLATPKAVKAAVELANGKYTAQDATTTAQAQGIQLSNATN    | 520 |
|  | QY | 408 | AGNOISIGNYRGWFNPGPSVYII   | 429 |
|  | DB | 521 | STSEMILA-----ATPKSVKAA  | 537 |

## RESULT 7

```

; US-11-052-554A-79
; Sequence 79, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL ME...
; TITLE OF INVENTION: PROTEINS OF THE...

```

```
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 79
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-79

Query Match      10.7%; Score 226; DB 7; Length 394;
Best Local Similarity 27.3%; Pred. No. 1.8e-06;
Matches 108; Conservative 63; Mismatches 177; Indels 48; Gaps 15;

QY 5 ILSAVLVGVTLSATLTSVAVKADDPDAQIASQ-----DSKINNLTAAQQAAQAQVNTIQ 59
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 16 ILSAVL--HVILPAALIWSS-----FDENIEASAGGGGSSIDAVWVDSGAVVEQTKRMQ 68
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 60 GQVSALQTQQAELQAEORLEAQSATLGGQIOTLSSKIVARNESLKQ-----QARSAQKS 114
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 69 SQESS-----AKRSDQKQKQQA-----BELREKQAAEQERLKQLEKERLAAQEQKK 118
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 115 NAATSYVINAINSKSVSDAINRVSAIREVVSANERKMLQQEODKAAVEQKQOENQAAINT 174
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 119 QAEAAKQAEKQKQAEAAKAAADAKAKAEADDKAAEAAKAAADAKKAAEAAKAA 178
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 175 VAANQETIAQNTNALNTQQAQLEAAQLNLQ-----AELTTAQDQKATLVAQKAAAEAA--R 229
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 179 AEAQKKAEEAAALUKKAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAA 235
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 230 QAAAAQAAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAA 289
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 236 KAAAEKAAADKKA--AKAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAA 287
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 290 EQSAQAQVNSDQESTTATEAQ--PSASSASTAVTANTSSANTYPAGQCTWGVKSLAP-- 346
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 288 KGNASPAAGSGNTKNGASGADINNYAGQIKSAIESKFYDASSYAGKCTCTLEIK-LAPDG 346
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 347 WYGNWNGGOWA--ASAAAAGYRVGSTPSAGAVAVW 381
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 347 MLDIKPEGGDPALCOALAAAKIAKIPKPPSQAVY 382
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
US-10-873-528-109
; Sequence 109, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129W0
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 109
; LENGTH: 1236
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-109

; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 79
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-79

Query Match      10.7%; Score 226; DB 7; Length 394;
Best Local Similarity 27.3%; Pred. No. 1.8e-06;
Matches 108; Conservative 63; Mismatches 177; Indels 48; Gaps 15;

QY 5 ILSAVLVGVTLSATLTSVAVKADDPDAQIASQ-----DSKINNLTAAQQAAQAQVNTIQ 59
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 16 ILSAVL--HVILPAALIWSS-----FDENIEASAGGGGSSIDAVWVDSGAVVEQTKRMQ 68
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 60 GQVSALQTQQAELQAEORLEAQSATLGGQIOTLSSKIVARNESLKQ-----QARSAQKS 114
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 69 SQESS-----AKRSDQKQKQQA-----BELREKQAAEQERLKQLEKERLAAQEQKK 118
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 115 NAATSYVINAINSKSVSDAINRVSAIREVVSANERKMLQQEODKAAVEQKQOENQAAINT 174
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 119 QAEAAKQAEKQKQAEAAKAAADAKAKAEADDKAAEAAKAAADAKKAAEAAKAA 178
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 175 VAANQETIAQNTNALNTQQAQLEAAQLNLQ-----AELTTAQDQKATLVAQKAAAEAA--R 229
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 179 AEAQKKAEEAAALUKKAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAA 235
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 230 QAAAAQAAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAA 289
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 236 KAAAEKAAADKKA--AKAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAAEAA 287
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 290 EQSAQAQVNSDQESTTATEAQ--PSASSASTAVTANTSSANTYPAGQCTWGVKSLAP-- 346
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 288 KGNASPAAGSGNTKNGASGADINNYAGQIKSAIESKFYDASSYAGKCTCTLEIK-LAPDG 346
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 347 WYGNWNGGOWA--ASAAAAGYRVGSTPSAGAVAVW 381
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 347 MLDIKPEGGDPALCOALAAAKIAKIPKPPSQAVY 382
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
US-10-793-626-1682
; Sequence 1682, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1682
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1682

Query Match      10.2%; Score 215; DB 6; Length 149;
Best Local Similarity 39.7%; Pred. No. 2.6e-06;
Matches 50; Conservative 19; Mismatches 43; Indels 14; Gaps 5;

QY 286 GDSTEQSAQAQVNSDQESTTATEAQPSASSASTAVVTANTSSANTYPAGQCTWGV-KSL 344
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 345 APWVGYNGNGQWAAASAAAAGYRVGSTPSAGAVWNDGGYGHVAYVTGQV-GQIQVQ 403
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 68 GKGIGTWNANSWATAQAAGFTVNTPPEGAINMQSEGAFGHVAFVESVNDGSIITVS 127
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 404 EANYAG 409
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 128 EMNYDG 133
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

|  |     |  |                                      |     |
|--|-----|--|--------------------------------------|-----|
| Qy   | 281 | ----   | QAAGTGDSTGCSAAQAVNNSDQESTTATEAQ----- | 311 |
| Dd   | 67  | FPGQVINVGSASQNTSSNTSSSASHTVVAGESLNI                          | IANKYGVSDALMQAHLNGYL                 | 126 |
| Qy   | 312 | -----PSASSASTAVVTANTSSA-----                                 | NTYPAGQCTWGV-----KSLAPWV             | 348 |
| Dd   | 127 | IMPNQILITPVGSGSGCGGTATQTSGNYTSPSFNHQNL                       | YTEGQCTWVFDKRSGAKPI                  | 186 |
| Qy   | 349 | GNTWNGGGWAASAAAAGYRVGSTPPSAGAVANNDDGYGHVAYTVTGQG-QGIQVQEANY  | 407                                  |     |
| Dd   | 187 | STYWSDAKYWASNAANDGYQVDNTPSVGAIMQSTPGPYGHVAYVERINGDGSILISEMNY | 246                                  |     |
| Qy   | 408 | AGNQSIGNYR   | 417                                  |     |
| Dd   | 247 | ANGPNMNRYR   | 256                                  |     |
| <br>RESULT 12<br>US-11-052-554A-211  |     |  |                                      |     |
| ; Sequence 211, Application US/11052554A                                   |     |  |                                      |     |
| ; Publication No. US20050288866A1  |     |  |                                      |     |
| ; GENERAL INFORMATION:   |     |  |                                      |     |
| ; APPLICANT: Sachdeva, et al.  |     |  |                                      |     |
| ; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL                    |     |  |                                      |     |
| ; FILE REFERENCE: 30853/40359A   |     |  |                                      |     |
| ; CURRENT APPLICATION NUMBER: US/11/052,554A                               |     |  |                                      |     |
| ; CURRENT FILING DATE: 2005-02-07  |     |  |                                      |     |
| ; PRIOR APPLICATION NUMBER: US 60/589,227                                  |     |  |                                      |     |
| ; PRIOR FILING DATE: 2004-07-20  |     |  |                                      |     |
| ; PRIOR APPLICATION NUMBER: IN 173/DEL/2004                                |     |  |                                      |     |
| ; PRIOR FILING DATE: 2004-02-06  |     |  |                                      |     |
| ; NUMBER OF SEQ ID NOS: 763  |     |  |                                      |     |
| ; SOFTWARE: Patent in version 3.3  |     |  |                                      |     |
| ; SEQ ID NO 211  |     |  |                                      |     |
| ; LENGTH: 1562   |     |  |                                      |     |
| ; TYPE: PRT  |     |  |                                      |     |
| ; ORGANISM: Streptococcus mutans UA159                                     |     |  |                                      |     |
| US-11-052-554A-211   |     |  |                                      |     |
| <br>Query Match            9.7%; Score 205; DB 7; Length 1562;             |     |  |                                      |     |
| Best Local Similarity   25.9%; Pred. No. 0.00014;                          |     |  |                                      |     |
| Matches     85; Conservative   62; Mismatches 113; Indels   68; Gaps   12; |     |  |                                      |     |
| Qy   | 28  | DFDAQIASODSKINLLTAQQOAAQAVNTIQGVSA                           | LQTCQAELOENQBLEAQSAT--               | 85  |
| Dd   | 151 | DOYKSDVAHEAEVAKIKAKNQTKEQ-----                               | YEKDMAAHKAEEVERINAANAASK             | 200 |
| Qy   | 86  | -----LGQOIQTLS--KIVARNESLUQARSA-                             | -----OKSNAA--TSYNAILI--              | 125 |
| Dd   | 201 | TAYEAKLAQYQADLAADVOKTNANQRAYQALAA                            | YQAEELKRVQENAAAKAYDTAVAAN            | 260 |
| Qy   | 126 | NKSVSDDAINRVSAIREVVSANEKMLOQOEQD                             | KAAVEKQOEHQAAINTVAANOETIAQN          | 185 |
| Dd   | 261 | NAKNT-----ELAAANEERKRNRATAKAEYET                             | KLQAYQAELEKRV-----QE                 | 301 |
| Qy   | 186 | TNALNTQQAQLRAAOLNLQAEILTADQOKATLV                            | AQKAAABEAFAQAAAAAKAAAE               | 245 |
| Dd   | 302 | ANAA-----EADYQAKLTAYOTELARVOKAN                              | ADAKAAEYAAVAAANNKNAALTAE             | 352 |
| Qy   | 246 | AKALQEQAAQAAQ-----AANNNTQATDAS                               | DOQAAAAADNT--QAAQTGDSTEOSAAQAVNNS    | 300 |
| Dd   | 353 | NTAIKORNENAKATYERALKQYEDLDLAAVK                              | KANAANEADYQAKLTAYCTELARVOKAN--       | 410 |
| Qy   | 301 | QDSETTTATEAQPSSASTAVVTANTSS                                  | 328                                  |     |
| Dd   | 411 | -ADAKAAEYAAVAAANNAALTAENTA                                   | 437                                  |     |
| <br>RESULT 13<br>US-10-857-780-23  |     |  |                                      |     |
| Sequence 23, Application US/10857780                                       |     |  |                                      |     |

```
; Publication No. US20050272043A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: BRAUN, ANDREAS
; APPLICANT: KAMMERER, STEFAN M.
; APPLICANT: NELSON, MATTHEW ROBERTS
; APPLICANT: RENELAND, RIKARD HENRY
; APPLICANT: HOYAL-WRIGHTSON, CAROLYN R.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
; FILE REFERENCE: SEQ-4069-CP
; CURRENT APPLICATION NUMBER: US/10/857,780
; PRIOR FILING DATE: 2004-05-28
; PRIOR APPLICATION NUMBER: 10/723,681
; PRIOR FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: 60/490,234
; PRIOR FILING DATE: 2003-07-24
; PRIOR APPLICATION NUMBER: 60/525,239
; PRIOR FILING DATE: 2003-11-25
; NUMBER OF SEQ ID NOS: 4962
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 2101
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-857-780-23

Query Match          9.4%; Score 198.5; DB 6; Length 2101;
Best Local Similarity 22.6%; Pred. No. 0.00046;
Matches 96; Conservative 62; Mismatches 180; Indels 87; Gaps 14;

QY 2 KKRILSAVLVSGVTLSSATLSA-----VKADDDFAQIASODSKI-----NNLTAAQQA 51
DB 405 KGEVLGDLVLQLETLKQEAATLAANNLTQLOARVEMLETERGQGEAKLLAERGHFEKQOL 464
QY 52 QAVNTIQGVSAQTQQAEL-----QAENORLEAQSATLGGQIQTLSSKIVARNE-----SL 104
DB 465 SSLITDLOSSINLSQAKLELQSAQAHGARTAQVASILTSLTILNATIQQDQELAGL 524
QY 105 KQARSQAQKNAATSYINAINSKSVDAINRVSAIREVSVANERKMLQQEQEDKAAVQK 164
DB 525 KQAKKEQAQLAQT-----LQQEQASQGLRHQVEQLSSSLKQKEQLKEVAEK 573
QY 165 Q-----QENQAAINTVAANOETIAQNTALNTQQAOLE-----AAQLNL----- 203
DB 574 QEATQDHAQQLATAAEEREASLRERDAALKQLEALEKEKAKEKLELQOOLQVANEARDS 633
QY 204 -QAEILTQAQDQKATL---VAQKAAAEAEARQAAAAQAAAEAKAAAEAKALQEQAAQAAA 259
DB 634 AQTSVTQAQREKAELSRKVEELQACVETARQEQHEAQVAELQELQLRSEQQKATEKERV 693
QY 260 ANNTQ-----ATDASDQAAAAADNTQAAATQAAQTGDSSTEQSAA 294
DB 694 AQEKDQLQOLQALKESLKVTKGSLSEEEKRRADALEEQQRCSIBLK-AETRSVLVEQHKR 752
QY 295 QAVN-NSDOESTTATEAQ-----PSASSASTAVV---TANTSSANTYPAGCTGWKSLAP 346
DB 753 ERKELEERAGRKGLLEARLLQIGEAHQAEIVLRRELAEMAAQHTASECEQLKVEAA 812
QY 347 VVGNY 351
DB 813 WRDGY 817

RESULT 14
US-10-793-626-3154
; Sequence 3154, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
```

```
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3154
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-3154

Query Match          9.3%; Score 196.5; DB 6; Length 1095;
Best Local Similarity 22.0%; Pred. No. 0.00028;
Matches 96; Conservative 63; Mismatches 177; Indels 101; Gaps 12;

QY 13 GVTLSATTLSAVKADDDFAQIASQDSKINNLTAAQQAQAQVNTIQGVSAQTQQAEL 72
DB 340 GMTTDTANNYSKKREAE-ELQKAAQIINNGDATEQQITNETNRVNQAINAKANDL 398
QY 73 QAENORLEAQSATLGGQIQTLSSKIVA-RNESLKKQARSQAQKNAATSYINAINIINS---- 127
DB 399 RADKSOLENAYNQLIQNVDTNGKKPASIQQYQAARQAIETQYNNAKSEAHQILENSPSV 458
QY 128 KSVSDAINRVSAIREVSVANERKMLQQEQEDKAAVEKQKQENQA-----AINT 174
DB 459 NEVAQALQKVEAVQKVNDAIHMQLQKNKNSALVTAKVQLQQAQVNDPFLTGMTODSINN 518
QY 175 VAA-----NOETIAQNTNA-----LNTQQAQLEAAQNLNQ-----AELTTA 210
DB 519 YVAKNEAQAIRNAEAVINNGDATAKQISDEKSKVEQALAHNDKAKQOLTADTTTELQTA 578
QY 211 QDQ-----KATLVAQKAAAEAEARQAAAAQAAAEAKAAAEAKALQEQAAQAAA 259
DB 579 VOQLNRGRDGNKKPRSNAYNKAIQSLQETQITSKADNANAVIQPIRTVQEVNNALQOV 638
QY 260 ANNTQATDASDQAAAAADN-----TQAQQTGDSSTEQS----- 292
DB 639 NQLNQQLTEAQLQPLSNNDALKAARLNLENKINQTVTQDGMTQOOSIEAYQNAKRVQ 698
QY 293 ----AAQAVNNSD-----QESTTATEAQPSSASSASTAVVTANTSSANTYPA 334
DB 699 ESNTALALINNGDADEQQITTEDRVNQTTNLITQAINGLTVNKEPLETEAKTALQNNDQ 758
QY 335 GQCTGWKSLAPWVGNV 351
DB 759 VPSTDGMTQQS--VANY 773

RESULT 15
US-10-485-517-212
; Sequence 212, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynex Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 212
; LENGTH: 1448
; TYPE: PRT
```

ORGANISM: Staphylococcus aureus  
US-10-485-517-212

Query Match 9.1%; Score 192.5; DB 6; Length 1448;  
Best Local Similarity 25.0%; Pred.No. 0.00066;  
Matches 92; Conservative 60; Mismatches 147; Indels 69; Gaps 16;

|    |     |  |
|----|-----|--|
| Qy | 7   | SAVLVSGVTLSSTATTLSAYK-----ADDFDAQIASQDSKINNLTAAQQQAAQAV-----55   |
| Db | 20  | NAKTTNEATIAITPDANVKPAKQAIADKVAQETAIDGNNGSTTTEKAAAKQVQTEK 79      |
| Qy | 56  | NTIQGVSAIQTQAELOAENQ-----RLEA-QSAT-----LCQOI 90                  |
| Db | 80  | TTADAADAHAHT-NAEVEAAKAAIAKIEAIOFATTTKDNKEAIAIKANERKTAIAQTQ 138   |
| Qy | 91  | QTLSSKIVARNESLKKQARSACKSNAATSYINAIINSKS-----VSDAINRVSAIRE 142    |
| Db | 139 | DITREEIAAANDVDNAVTOANSNIEAANSQNDVDQAKTTGENSIDQVPTVKKATARN 198    |
| Qy | 143 | VVSA--NEKMLQQEQODKAAVEQKQ-QENQAAINTVAANQETIAQNTNALNTQQAQLEAA 199 |
| Db | 199 | EITAILNNKLQEIQTATPDATDEEKQAADAEANTENGKANQAISAATTN-----AQVDEA 252 |
| Qy | 200 | QLNLQAEIQTADQKATLVAQKAAAEAAQAAAAQAAA---EAKAAAEAK--ALQEQAA 254    |
| Db | 253 | KANAEAAINAVTPK-----VVKQAQAKDEIDQLQATQTNVINNDQNATTEKEAAIQQLAT 308 |
| Qy | 255 | QAQAAANNNTQATDAS--DQQAADNT-QAAQTGDSFGSAAQAVNNSDQESTTATEAQ 311    |
| Db | 309 | AVTDKNNITAAATDONGVDQAKDAGKNSIQSTQPATAVKSNK---NDVDQAVTTQONQAI 365 |
| Qy | 312 | PSASSAST 319   |
| Db | 366 | DNTTGATT 373   |

Search completed: February 15, 2006, 18:35:26  
Job time : 9.32046 secs

**THIS PAGE BLANK (USPTO)**